

# When Elites Divide: Ethnic Party Segmentation Across the World\*

John Gerring †

Connor T. Jerzak‡

Erzen Öncel §

## Abstract

Scholars agree that the nature of ethnic cleavages hinges not on diversity per se but on the degree to which political elites reproduce social boundaries. Yet, we still lack a replicable, continuous, and cross-national metric of elite ethnic segregation of the party system. This paper proposes a new measure that treats the distribution of ethnic groups across legislative party delegations as an optimal-transport problem. By comparing the observed joint distribution of groups and parties with a counterfactual of perfect integration, we derive a segregation index bounded in [0, 1] that is both decomposable and comparable across time, chambers, and countries. We implement the index on an original dataset covering 219 parliamentary delegations (1990–2022), assembled from the Global Leadership Project and supplemented by a provenance-preserving coding AI search protocol that harvests and cites public-source evidence to infer missing party labels (expanding coverage by 20.8% (12,898 leader cases)). Validation against existing binary and survey-based measures shows that our cleavage metrics capture known cases while revealing substantial within-country and within-party variation that is missed by existing approaches. Substantively, we find patterns consistent with a simple demand-and-supply model of ethnic representation. The index invites new tests of classic theories of ethnic politics, affords fine-grained diagnostics for institutional engineering, and is readily extensible to other ascriptive cleavages. We make the 770,000 distinct records on political elites open source.

**Keywords:** Descriptive representation; Political parties; Social groups

---

\*Authors are listed in alphabetical order. Preliminary and incomplete. For more information about dataset access, see [GlobalLeadershipProject.net/cleavage](http://GlobalLeadershipProject.net/cleavage).

†Professor, Department of Government, University of Texas at Austin. Email: [jgerring@austin.utexas.edu](mailto:jgerring@austin.utexas.edu)  
URL: [liberalarts.utexas.edu/government/faculty/jg29775](http://liberalarts.utexas.edu/government/faculty/jg29775)

‡Assistant Professor, Department of Government, University of Texas at Austin. Email: [connor.jerzak@austin.utexas.edu](mailto:connor.jerzak@austin.utexas.edu)  
URL: [ConnorJerzak.com](http://ConnorJerzak.com)

§Assistant Professor, Department of International Relations, Özyegin University. Email: [erzen.oncel@ozyegin.edu.tr](mailto:erzen.oncel@ozyegin.edu.tr)  
URL: [ozyegin.edu.tr/en/faculty/erzenoncel](http://ozyegin.edu.tr/en/faculty/erzenoncel)

# Contents

<b>1</b>	<b>Measuring Cleavages with Optimal Transport</b>	<b>3</b>
1.1	Ethnic Cleavages: Concept & Intuition . . . . .	4
1.2	A Legislative Segmentation Index . . . . .	5
1.3	A Party Segmentation Index . . . . .	7
1.4	A Group Segmentation Index . . . . .	8
<b>2</b>	<b>Data Collection and Coding</b>	<b>8</b>
2.1	Expert Coding . . . . .	9
<b>3</b>	<b>Ethnic Cleavages Across the World</b>	<b>9</b>
3.1	Party and Group Segmentation . . . . .	12
3.2	Convergent Validity: Validation Against Existing Measures . . . . .	15
<b>4</b>	<b>Explanations</b>	<b>15</b>
4.1	Simulations . . . . .	15
4.2	Observational Evidence . . . . .	16
<b>5</b>	<b>Do Cleavages Matter?</b>	<b>18</b>
<b>6</b>	<b>Conclusion</b>	<b>19</b>
<b>7</b>	<b>Appendix A</b>	<b>23</b>
7.1	Additional Discussion . . . . .	23
7.2	Additional Analyses . . . . .	23
7.3	Results, Expert Only . . . . .	25
7.4	Appendix B: Simulations . . . . .	28
<b>8</b>	<b>Appendix C</b>	<b>30</b>
<b>9</b>	<b>Appendix D</b>	<b>51</b>
9.1	AI Agent Details . . . . .	51

## Introduction

An enormous literature wrestles with challenges to governance and development posed by ethnic diversity (Alesina and La Ferrara, 2005; Easterly and Levine, 1997). Although the concern is well-founded, one must also appreciate that ethnic differences by themselves are neither unusual nor inherently problematic for governance. Every society is diverse in some respects, and in most instances this diversity does not prevent cooperation; it may even promote better solutions (Page, 2008). Diversity becomes consequential when latent differences are politicized (Fraenkel and Crawley, 2025; Posner, 2004). This is why it is important to distinguish ethnic differences (as measured, e.g., by ethnic fractionalization indices) from ethnic *cleavages*, our subject in this study.

Of particular concern to political scientists is the prospect that cleavages of an ethnic nature might undermine stability and democracy. The first generation of studies offers a pessimistic view, according to which ethnic politics is an invitation to civil conflict (Geertz, 1963; Horowitz, 1985;

Rabushka and Shepsle, 1972; Rae and Taylor, 1970; Rustow, 1970). Some more recent studies confirm this pessimistic view (Houle, 2018).

Others are more optimistic, pointing out that the existence of ethnic cleavages does not necessarily convert elections into an ethnic census (Elischer, 2013; Fraenkel and Crawley, 2025). Moreover, the potentially damaging effects of ethnic cleavages may be mitigated by political institutions such as federalism, electoral rules, or power-sharing agreements (Chandra, 2005; Lijphart, 1977; Reilly, 2001) or by alliances with the private sector (Arriola, 2012). Indeed, the theory of consociationalism suggests that partisan cleavages based on ethnicity can be successfully mediated by appropriate institutions (Lijphart, 1969). Evidently, the existence of ethnic parties does not doom prospects for democracy (Birnir, 2006; Chandra, 2007; Ishiyama, 2009).

Even so, where voters choose leaders on the basis of who they are rather than what they stand for, this may undermine mechanisms of accountability essential for good governance. Accordingly, ethnic politics, associated with clientelistic policymaking and particularistic goods, is commonly counterposed to class politics, associated with programmatic policymaking and the provision of public goods (Chandra, 2007; Dixit and Londregan, 1996; Kitschelt, 2007; Lemarchand, 1972).

The role of ethnic identities in governance remains a point of contention. Issues are difficult to resolve because there is no generally recognized metric of ethnic cleavages. What makes one polity more ethnically divided than another, one party more ethnically defined than another, or one ethnic group more politicized than another?

In this study, we offer a new approach to the measurement of ethnic cleavages. This approach has several distinctive characteristics that, collectively, set it apart from extant work. It treats ethnicity as a political feature rather than a demographic fact. It is centered on elites rather than voters. It approaches cleavages as matters of degree rather than of kind. And it is applicable to any setting where the ethnic identity of parliamentarians can be ascertained.

Our key empirical indicator is the ethnic composition of party delegations. We describe delegations as *integrated* if there are few ethnic differences across parties and *segmented* if each party represents a different ethnic group and these groups are similar in size. To provide a precise measure of this fuzzy concept, we adopt an algorithm derived from optimal transport theory (Lott and Villani, 2009). Applied to party delegations in a legislature, this decomposable approach generates an *legislative segmentation index*. A low score on this index indicates a party system where ethnicity plays little role: either there are no sizeable ethnic groups (MPs are ethnically homogeneous) or parties have balanced delegations. A high score indicates a party system defined by ethnicity. Adaptations of optimal transport theory apply the same principles to political parties and ethnic groups. A *party segmentation index* measures the degree to which individual parties are defined by ethnicity. An *ethnic segmentation index* measures the degree to which individual ethnic groups are represented by the same party.

After introducing these measures, we turn to the task of data collection. Evidently, mathematically elegant indices are useful only if they can be successfully applied to topics of theoretical interest. To meet this challenge, we propose an approach to coding the ethnicity of legislators around the world that leverages hand-coding along with a new set of LLM-based workflows.

This protocol allows for a global dataset measuring ethnic cleavages across legislatures in [] countries in the contemporary era. In the third section of the paper we introduce the data and compare it with other measures of ethnic voting and ethnic parties.

In the fourth section of the paper we take up the task of explanation. Why are ethnic cleavages more marked in some polities than in others? We argue that this is largely a product of the

composition of ethnic groups and of political parties. Greater fractionalization generally translates into stronger cleavages.

The final section of the paper explores the potential importance of our measure of ethnic cleavages. We show that our legislative segmentation index is more closely related to outcomes such as democracy, corruption, and conflict than other measures of ethnic politics. This suggests that elite-level cleavages may be more consequential than mass-level cleavages.

## 1 Measuring Cleavages with Optimal Transport

Cleavage structures have been a preoccupation of political science and sociology for the better part of a century. Originally centered on social class (Alford, 1962; Converse, 1958), researchers soon broadened their purview to include other aspects of status and identity such as language, religion, and region (Lipset and Rokkan, 1967).<sup>1</sup> Contemporary studies of ethnic politics follow in this venerable tradition, with ethnicity as the umbrella term encompassing related factors such as religion, language, and region.

Two principal approaches may be identified from the literature, centered respectively on ethnic groups and political parties (Huber, 2012). One focuses on the degree to which each ethnic group consolidates behind a single party; the other focuses on the degree to which each party represents a unique ethnic group. To grasp the distinction, consider a setting with four ethnic groups ( $A$ ,  $B$ ,  $C$ ,  $D$ ) and two parties ( $P_1$ ,  $P_2$ ). If all voters from groups  $A$  and  $B$  support  $P_1$ , and all voters from groups  $C$  and  $D$  support  $P_2$ , *ethnic voting* is maximal and *ethnic parties* is minimal.

Beyond this conceptual disagreement lie empirical differences. When attempting to ascertain ethnic voting, some researchers focus on vote shares across constituencies. This sort of data is usually plentiful but raises problems of ecological inference since the characteristics of individual voters must be inferred from the characteristics of constituencies (obtained from census data). Other researchers rely on public opinion surveys. This provides individual-level data but is limited to contexts where national surveying is common and reliable, and where questionnaires include vote-choice (or party membership) and ethnic identity. Both data sources must contend with the specificity of census and survey data and the highly contextual dynamics of ethnicity, which may impair comparability across settings and even across surveys or censuses in the same setting (if the coding of ethnicity varies).<sup>2</sup>

In light of these obstacles, it is not surprising that most work on ethnic politics is limited to individual countries or regions. Recent studies center on Latin America (Madrid, 2012), Southeast Asia (Liu and Ricks, 2022; Reilly, 2021), the OECD (Hamza and Maeda, 2025), and Africa, where the topic is ubiquitous (Huber, 2012; Ishiyama, 2012). A few studies are more extensive, notably Houle, Park, and Kenny (2019), which covers sixty-five countries with data from the World Values Survey, and Fraenkel and Crawley (2025), which covers 132 countries with a wide variety of surveys.

A rather different approach categorizes individual parties as “ethnic” or “nonethnic” based on a variety of characteristics such as party name, rhetoric, policies, leadership, constituency, and expert judgments (Chandra, 2011). This corresponds to the party-centered approach to ethnic politics introduced above. Although most studies in this vein are limited to a single country or region, a few are more wide-ranging. Ishiyama (2009) codes ethnic parties in the developing world during the 1990s. Strijbis and Kotnarowski (2015, p. 463) code ethnic parties at various points over the

<sup>1</sup>For a recent effort, see Marks et al. (2023).

<sup>2</sup>For discussion of these and other difficulties encountered by survey-based research see Fraenkel and Crawley (2025, pp. 15–18).

past two decades in a handful of European countries, along with Canada and Australia. Van Cott (2007) codes indigenous parties in South America. Lublin (2014, App II) codes ethnoregional parties (considered jointly) in 80+ countries, observed at some point between 1990 and 2012.

While these studies offer broader coverage than the typical study of ethnic voting, it is important to appreciate the considerable loss of information that arises when parties are reduced to a single binary code — ethnic or nonethnic. This simplification complicates inferences one might draw about other parties (relegated to a large residual category) and the party system as a whole. Moreover, the complexity of the coding criteria means that raters must juggle a variety of dimensions that do not always point in the same direction. Since different studies of ethnic parties invoke different coding criteria, results differ somewhat across studies and are not easy to replicate.

Against this backdrop, our approach has five distinguishing features. First, it centers on elites (representatives) rather than masses (voters).<sup>3</sup> Second, it treats cleavages as matters of degree rather than of kind. Third, it assigns scores to individual parties (mirroring the party-centered approach) and ethnic groups (mirroring the group-centered approach) as well as polities (the overall cleavage). Fourth, it is applicable to any setting—local, regional, or national—where the party and ethnic identity of MPs can be ascertained, raising the possibility of a truly comprehensive analysis on a global scale.

## 1.1 Ethnic Cleavages: Concept & Intuition

In modern contexts, ethnic cleavages of any political significance are usually manifested in political parties. Where a cleavage is politically salient, parties will presumably be differentiated by ethnicity. Since the legislature is the preeminent representative body, it is natural to look to legislatures if we wish to understand the character of a party at elite levels. Helpfully, legislative parties are sizeable enough to offer a basis for judgment.<sup>4</sup>

Our guiding assumption is that a party’s ethnic orientation is reflected in the descriptive characteristics of its leadership. If the party’s mission is to represent the interests of a particular ethnic group, that group is likely to dominate its parliamentary delegation. If its mission is to represent a variety of different ethnic groups, this will be reflected in a multi-ethnic delegation.<sup>5</sup>

This does not mean that parties with an ethnic base must trumpet their ethnic character. In Africa, most parties are formally non-ethnic; yet, it is an open secret that many are vehicles for particular ethnic groups (Berman, Eyoh, and Kymlicka, 2004). The same was true of the U.S. Republican Party through most of its history. Temperance, education, Sabbatarianism, anti-bossism, and other “reform” issues were calculated to please Protestant constituencies despite the displeasure they caused Catholics. Although the party did not proclaim itself Protestant, it was responding to a constituency with roots in Protestant communities outside the South, a feature reflected in its electoral base and in its staunchly Protestant leadership (Gould, 2007; Layman, 2001; Silbey and Bogue, 1978). Accordingly, we consider the ethnic identity of a party’s legislative delegation to be strong evidence of its ethnic (or non-ethnic) orientation.

---

<sup>3</sup>As such, we sidestep complicated questions about voter motivation that are central to the literature on ethnic voting (Adida et al., 2017).

<sup>4</sup>By contrast, top leadership positions—presidents, prime ministers, party leaders—are encapsulated in a single office, which by construction can be occupied only by a single individual and thus allows no basis for distinguishing an ethnic party from a multi-ethnic one.

<sup>5</sup>It is of course possible for parties to dissemble. Leaders may nominate members of social groups they have no intention of representing (substantively), and for a while, they may be successful in this ruse. However, it is unlikely they will be successful for very long.

To operationalize the concept of an ethnic cleavage, we focus on the degree of alignment between ethnicity and party delegations. Where cleavages are extreme, each party represents one and only one ethnic group, a setting we describe as ethnic *segmentation*. In a fully *integrated* system, ethnic groups are distributed across parties in the same proportion as across the legislature at-large, or there are no discernible ethnic distinctions among MPs at all. Here, ethnicity appears to play no role in party cleavages. Our approach thus melds group- and party-centered approaches to the measurement of cleavages.<sup>6</sup>

To make these intuitions more concrete, several stylized scenarios are illustrated in Table 1. In each scenario, a legislature is divided between three parties of varying sizes and composition. In Legislature I, only one ethnic group (*A*) gains entrance into the legislature.<sup>7</sup> In Legislature II, three ethnic groups (*A-C*) are equally represented across three parties (*P1-P3*). In both of these scenarios, there is perfect integration. In Legislature VI, each ethnic group is represented by a different political party. This exemplifies the extreme case of segmentation, where all parties are ethnic parties and all ethnic groups are affiliated with a single party.

Note that Legislatures II and VI provide exactly the same parliamentary representation for the three ethnic groups. With respect to descriptive representation at the parliamentary level, one might conclude that these scenarios are equivalent. However, the principle of representation is radically different—cross-party (integrated) in Legislature II and pillarized (segmented) in Legislature VI. We suspect this has important ramifications.

**Table 1:** Illustrative scenarios.

Legislatures (N)	I (12)			II (12)			III (12)			IV (12)			V (14)			VI (12)		
Ethnic groups (N)	A (12)			A (4)	B (4)	C (4)	A (8)	B (4)	A (10)	B (1)	C (1)	A (8)	B (4)	C (2)	A (4)	B (4)	C (4)	
Parties (N)	P1 (4)	P2 (4)	P3 (4)	P1 (4)	P2 (4)	P3 (4)	P1 (6)	P2 (4)	P3 (2)	P1 (5)	P2 (5)	P3 (2)	P1 (4)	P2 (4)	P3 (6)	P1 (4)	P2 (4)	P3 (4)
MPs	a a a a	a a a a	a b a a	a b c d	a b c d	a b c d	a a a a	a a a a	a b a a	a a a a	a a a a	b b b a	a a a a	b b b b	a a a a	b b b b	c c c c	
Legislature Index	0			0			0.22			0.28			0.44			0.66		

## 1.2 A Legislative Segmentation Index

The extreme scenarios laid out in Legislatures I, II, and VI mark two ends of a continuum. In between lie an infinite variety of intermediate scenarios, a few of which are illustrated by Legislatures III-V.

To provide a sensitive metric that expresses all possible points on this continuum of integration and segmentation, we focus on the magnitude of the transformations required for a legislature to reach perfect integration. This scale ranges from 0 (where no changes are required) to a value

<sup>6</sup>In this respect, we follow Hamza and Maeda (2025), though the latter is focused on ethnic voting rather than elites.

<sup>7</sup>This is a complicated feature to digest since all parties are ethnically homogeneous (exemplifying ethnic parties) but there is no differentiation across them (exemplifying a non-ethnic regime). Tellingly, Horowitz (1985) does not provide a definition for a non-ethnic party, a point noted by Elischer (2013).

Observed O			Deviation $\Delta = \mathbf{O} - \mathbf{R}$			Perfect integration R					
	P1	P2	P3		P1	P2	P3		P1	P2	P3
Group 1	24 80%	5 17%	1 3%	$\Sigma = 30$	+9	-5	-4		15 50%	10 33%	5 17%
	5 28%	10 56%	3 17%	$\Sigma = 18$	-4	+4	0		9 50%	6 33%	3 17%
Group 2	1 8%	5 42%	6 50%	$\Sigma = 12$	-5	+1	+4		6 50%	4 33%	2 17%
	$\Sigma = 30$	$\Sigma = 20$	$\Sigma = 10$						$\Sigma = 30$	$\Sigma = 20$	$\Sigma = 10$

**Mechanics.** Under a 0–1 cost for switching across parties within groups, an optimal transport plan would reassign  $5 + 4 + 4 + 4 + 1 = 18$  seats. Hence  $W = \frac{1}{2} \sum_{i,j} |o_{ij} - r_{ij}| = 18$ , total seats  $n = 60$ , and the cleavage index  $S = W/n = 0.30$ . Larger  $S$  means more reassignment is required to make each party mirror the chamber’s group composition.

**Figure 1:** LEFT: observed joint distribution of groups by party (**O**) with within-group percentages and margins. CENTER: signed deviations  $\Delta = \mathbf{O} - \mathbf{R}$  (red=excess, blue=deficit); headers shown above the deviation panel for readability. RIGHT: perfect-integration target (**R**) preserving row/column totals so each party mirrors the chamber’s group mix. An optimal reassignment plan (flows omitted here) moves  $W$  seats; normalizing  $W$  by  $n$  yields the cleavage index  $S$ .

that approaches 1 asymptotically (where nearly everything needs to change in order to achieve integration, as the number of groups gets large). We refer to this as an *legislative segmentation index*.

To put this notion into motion, we begin with the observed distribution of parties by ethnic group in a legislature. We then measure the distance between this observed arrangement and a hypothetical arrangement under perfect integration. In this ideal scenario, the marginal distribution of groups and parties is preserved. Within that constraint, there is a uniform allocation of members across groups.

To quantify distance from perfect integration, we employ techniques from optimal transport theory (Lott and Villani, 2009). This mathematical framework provides a useful method for measuring the distance between two probability distributions. It also melds the party- and ethnic-centered approaches to ethnic cleavages introduced at the outset.

In this index, we pose a simple question: What is the smallest fraction of legislative seats that would need to be reassigned across parties—while keeping each MP’s ethnicity fixed—to ensure that every party’s delegation mirrors the chamber’s overall ethnic makeup? We cast this as a discrete optimal-transport problem (ibid.). Each cell in the groups, $\times$ , parties table is a “location.” Leaving an MP in the same cell costs 0; moving an MP to any other cell costs 1. Because the “perfect-integration” target preserves each group’s total number of MPs (row sums), any optimal reassignment can be implemented *within* ethnic groups only; conceptually, changing an MP’s ethnicity carries an infinite penalty and is therefore ruled out. With this 0–1 cost structure, the transport objective equals the *total variation distance* (half the  $\ell_1$  difference) between the observed joint distribution and the integration target. The resulting score lies in  $[0, 1]$  and has a direct interpretation: it is the minimal *share of seats* that must be reallocated across parties (within groups) to achieve perfect integration.

To get a feel for this way of measuring cleavages, let us return to the examples sketched in Table

- Scenarios I and II exemplify perfect integration, so these legislatures receive a perfect score of 0 (no segmentation) in the bottom row.

Scenario VI exemplifies perfect segmentation. Nearly everything—but not quite everything—must change in order to achieve perfect integration across parties. Specifically, one would have to reassign all but one MP per party in order to achieve integration. This is why it is an asymptotic value. In cases of purely ethnic parties, the value of the segmentation index approaches 1 as the number of groups becomes larger.

Importantly, as the number of parties shrinks the expected segmentation of the legislature also shrinks, approaching zero in the case of a single-party system. Where  $N(\text{parties}) = 1$ , segmentation = 0. Likewise, as the number of ethnic groups shrinks, so does the expected segmentation score (see Appendix XXX). Where  $N(\text{ethnic groups}) = 1$ , segmentation = 0. These compositional effects play a crucial role in our explanatory framework, outlined in Section 4.

### 1.3 A Party Segmentation Index

The same logic may be extended to scoring for individual parties. Indeed, the score for a legislature is simply the weighted average of the segmentation scores of all the parties in the legislature.

Consider the example illustrated in Table 2. Here, we see a legislature with sixteen MPs, three parties, and two ethnic groups. The majority group,  $A$ , comprises 3/4 of the legislature while the minority group,  $B$ , comprises 1/4.

**Table 2:** Ethnic representation across parties. Groups  $A$  and  $B$  are represented in varying proportions.

Legislatures ( $N$ )		<b>I</b>	
Ethnic groups ( $N$ )		<b>A</b> (12)	<b>B</b> (4)
Parties ( $N$ )		<b>P1</b> (8)	<b>P2</b> (6)
<i>MPs</i>		<i>b</i>	<i>a</i>
		<i>b</i>	<i>a</i>
		<i>a</i>	<i>a</i>
<i>Party Index</i>		XXX	XXX
		XXX	XXX

Perfect integration for a party means that the distribution of ethnicities in its delegation mirrors the distribution of ethnicities in the legislature as a whole. This is the case for Party 1, which receives a score of 0, indicating perfect integration.

Where a party deviates, one must calculate the degree of deviation, which may be understood loosely as the magnitude of changes required to achieve perfect integration.

Although Party 2 is composed exclusively of one ethnic group,  $A$  happens to be the largest group in the legislature (by far). Accordingly, it requires only a small change in composition for Party 2 to mirror the legislature, rendering a modest ethnic party score. This exemplifies the situation of "ethno-nationalist" parties such as Fidesz in Hungary, United Russia in Russia, or the BJP in India (Rydgren, 2007).

Party 3 is also composed of a single ethnic group. However, this group comprises a small minority of the legislature. As such, its membership must be thoroughly transformed in order to mirror the ethnic characteristics of the legislature. This exemplifies the situation of "ethnic parties," introduced at the outset.

## 1.4 A Group Segmentation Index

The third elaboration of social cleavages focuses on the extent to which members of an ethnic group are concentrated in a single party or dispersed across multiple parties. To fix ideas, Table 3 explores three ethnic groups, each with three MPs. Group  $A$  is concentrated on one party. Group  $B$  is spread across two parties. And Group  $C$  is dispersed across all three parties.

**Table 3:** Ethnic representation across groups. Groups  $A$ ,  $B$ , and  $C$  shown seat-by-seat.

Legislatures (N)		I (9)		
Parties (N)		P1 (6)	P2 (2)	P3 (1)
MPs		<i>a</i>	<i>b</i>	<i>c</i>
		<i>a</i>	<i>c</i>	
		<i>a</i>		
		<i>b</i>		
		<i>b</i>		
		<i>c</i>		
Ethnic groups (N)		A (3)	B (3)	C (3)
Group Index		XXX	XXX	XXX

To provide a precise measure of cohesion, we again leverage optimal transport, now from the group perspective. We compute a cohesion distance that compares how that group's MPs are spread across parties to the chamber's overall distribution of seats by party; it is interpreted as the smallest share of that group's MPs who would need to change parties—while keeping ethnic identities fixed—for the group to be proportionally represented in every party.

## 2 Data Collection and Coding

Having laid out our approach to measuring ethnic cleavages across legislatures, parties, and groups, we turn to the empirical study of these matters. Data collection and coding are much more than an afterthought as they involve solving difficult challenges of conceptualization and measurement.

We begin with ethnicity, about which so much has been written (Abdelal et al., 2009; Marquardt and Herrera, 2015). For present purposes, this concept encompasses any ascriptive identity, i.e., any set of group characteristics understood as inherited, including customs, language, race,

region, and various combinations of the foregoing. We reserve the word "ethnic" for the most salient of these dimensions (or combination thereof) as understood in a particular society at a particular point in time. This recognizes the constructed nature of ethnicity as well as its capacity for change.

Since our intention is to assess how ethnic groups are represented, not whether they are represented (at all), our purview is limited to groups that gain some political representation at national levels. Very small ethnic groups or groups that face intense discrimination or are denied citizenship are likely to be excluded. This is in keeping with common understandings of the concept of a political *cleavage*, which refers to a relationship between political parties and constituencies that are allowed to participate and to attain public office. A racial political cleavage in the American South appears only when blacks were enfranchised, for example.

Likewise, we do not consider whether the representation of groups is proportional to their population. This is, of course, an important issue (Gerring, Jerzak, and Öncel, 2024), but it is orthogonal to the concept of cleavage structures. A party whose leaders are drawn exclusively from a single ethnic group is no more or less ethnic if that group is over- or under-represented.

## 2.1 Expert Coding

To code the ethnicity of MPs, we rely primarily on country experts enlisted for the Global Leadership Project (*ibid.*). Their judgments rest on cues about each MP drawn from names, birthplaces, and photos—often contained in parliamentary websites. Bear in mind that we are interested in how MPs represent themselves to their constituents. For present purposes, the ethnicity of Representative  $X$  is whatever  $X$  says or implies it to be—their public presentation of self (Goffman, 1959). We do not concern ourselves with whether  $X$  is really (authentically) Christian or Muslim, Croat or Serb.

Despite this caveat, we acknowledge that different judgments about how to define ethnic groups in a given country might lead to different conclusions. The same is true for extant codings of ethnic voting and ethnic parties (reviewed above), which depend upon prior judgments about what an ethnic group is and how it should be operationalized in a given context—often, a fraught exercise subject to ongoing revision (Csata, Hlatky, and Liu, 2021). We see no way around this conundrum. Readers may survey the decisions reached by our coders, listed in Appendix ??, and decide for themselves.

Expert coding is supplemented with a subset of high-confidence parties from an LLM-based search agent that analyzes Wikipedia and search engine results, and bases its prediction of the relevant missing value on the search content found. Median accuracy across countries is above 0.90. For more details, see Appendix XXX.

## 3 Ethnic Cleavages Across the World

Having outlined our protocol for data collection, we turn to a survey of ethnic cleavages across the world. Table 4 offers a list of all countries in our dataset ( $N=130$ ), ordered by the score on the legislative segmentation index. Since most countries are observed at several points in time (usually separated by an election) these observations are averaged to obtain a single estimate for each country.

In the upper tier of Table 4 are countries where party competition closely tracks ascriptive communities—often settings with institutionalized communal representation, pillarized party histories, and regionally concentrated electorates, e.g., Lebanon, Bosnia and Herzegovina, Malaysia, Belgium, India.

A broad middle comprises varied systems in which explicitly ethnic parties coexist with cross-

ethnic, programmatic, or catch-all organizations; this stratum includes several multilingual democracies and diverse federations where regional or linguistic cleavages are politically salient but not uniformly determinative, e.g., Spain, Canada, New Zealand, and parts of sub-Saharan Africa.

At the lower end cluster three sorts of cases: (i) non-competitive or hegemonic-party regimes that prevent the expression of ethnic cleavages (e.g., North Korea, Venezuela); (ii) relatively homogeneous polities where ethnic differences are muted (e.g., Japan, Norway) and (iii) systems whose principal conflicts are ideological rather than ascriptive (e.g., much of Western Europe).

**Table 4:** Index by country.

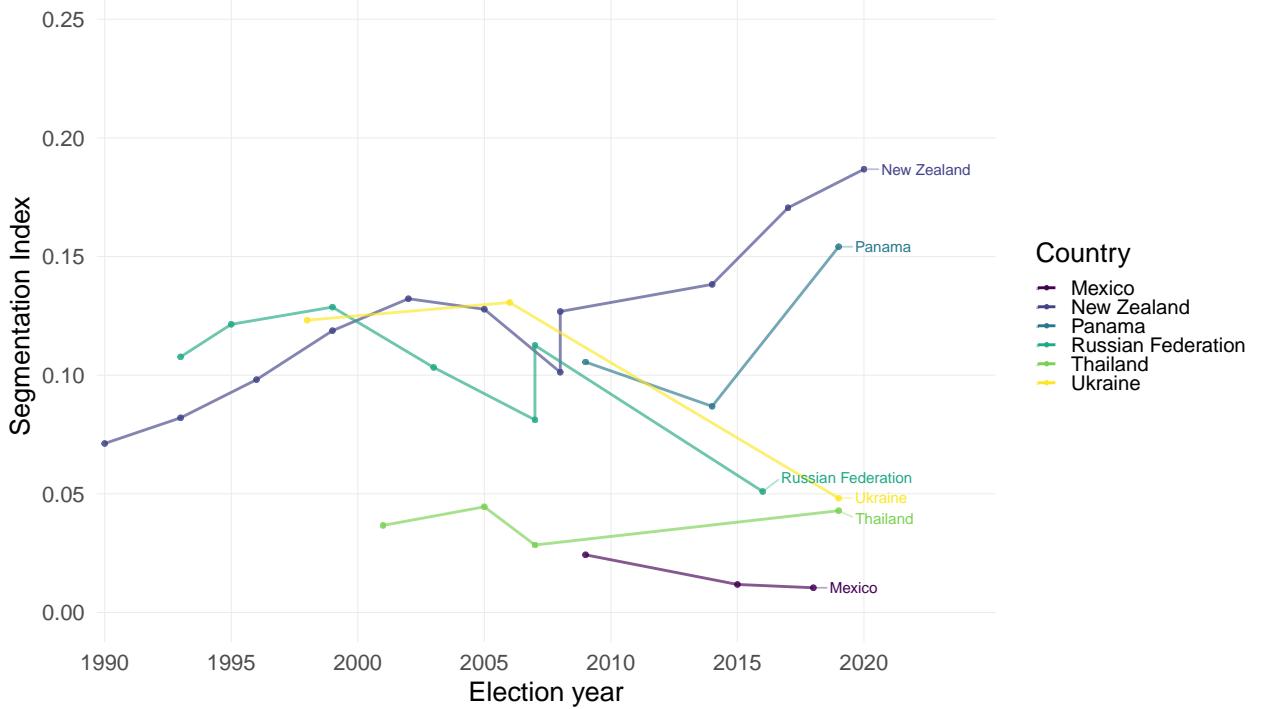
Country	Index	Country	Index	Country	Index
Congo	0.568 (1)	Ethiopia	0.130 (45)	Denmark	0.031 (89)
Bosnia & Herz.	0.566 (2)	El Salvador	0.128 (46)	Germany	0.027 (90)
Lebanon	0.540 (4)	Switzerland	0.126 (47)	Fiji	0.024 (91)
Malaysia	0.540 (4)	Iraq	0.123 (49)	Venezuela	0.023 (93)
India	0.461 (5)	New Zealand	0.123 (49)	Hungary	0.023 (93)
Belgium	0.425 (6)	Panama	0.116 (50)	Georgia	0.020 (94)
Solomon Is.	0.422 (7)	Slovakia	0.110 (52)	Chile	0.018 (95)
Kenya	0.370 (8)	Morocco	0.110 (52)	Kazakhstan	0.017 (96)
Pakistan	0.362 (9)	Turkey	0.108 (53)	Mexico	0.015 (98)
Mauritius	0.348 (10)	Peru	0.107 (56)	Costa Rica	0.015 (98)
Benin	0.344 (11)	Kosovo	0.107 (56)	Singapore	0.011 (102)
Latvia	0.341 (12)	Gabon	0.107 (56)	Cape Verde	0.011 (102)
Moldova	0.322 (13)	Paraguay	0.105 (57)	Ireland	0.011 (102)
Macedonia	0.320 (14)	Senegal	0.104 (58)	Honduras	0.011 (102)
Myanmar	0.318 (15)	Botswana	0.103 (59)	Cambodia	0.010 (106)
Afghanistan	0.317 (16)	Djibouti	0.102 (60)	Austria	0.010 (106)
Guyana	0.273 (17)	Russia	0.101 (62)	China	0.010 (106)
Burkina Faso	0.269 (18)	Ukraine	0.101 (62)	Albania	0.010 (106)
Israel	0.267 (19)	Finland	0.099 (63)	Armenia	0.008 (108)
Mali	0.256 (20)	Madagascar	0.095 (65)	Italy	0.008 (108)
Niger	0.254 (21)	Namibia	0.095 (65)	Egypt	0.007 (112)
C. Afr. Rep.	0.253 (22)	Rwanda	0.091 (66)	Azerbaijan	0.007 (112)
S. Africa	0.252 (23)	Dom. Rep.	0.089 (67)	Poland	0.007 (112)
Zambia	0.250 (24)	Canada	0.087 (68)	Greece	0.007 (112)
Guinea	0.241 (25)	Serbia	0.084 (70)	Uruguay	0.004 (114)
Montenegro	0.237 (26)	Lithuania	0.084 (70)	Belarus	0.004 (114)
Indonesia	0.218 (27)	Czech Rep.	0.083 (71)	Tajikistan	0.002 (116)
Ghana	0.217 (28)	Burundi	0.070 (72)	Bangladesh	0.002 (116)
Gambia	0.211 (30)	Nicaragua	0.067 (74)	Portugal	0.001 (118)
Trin. & Tob.	0.211 (30)	Netherlands	0.067 (74)	Tunisia	0.001 (118)
Guinea-Bissau	0.192 (31)	Zimbabwe	0.064 (75)	Viet Nam	0.000 (130)
Sierra Leone	0.188 (32)	Sweden	0.055 (76)	Cuba	0.000 (130)
Ecuador	0.187 (33)	Slovenia	0.054 (77)	Haiti	0.000 (130)
Bolivia	0.184 (34)	Croatia	0.053 (79)	Iceland	0.000 (130)
Spain	0.183 (35)	Brazil	0.053 (79)	Japan	0.000 (130)
Malawi	0.163 (36)	Mongolia	0.052 (80)	Jordan	0.000 (130)
Iran	0.155 (37)	Uzbekistan	0.051 (81)	Lesotho	0.000 (130)
Sri Lanka	0.152 (38)	Philippines	0.049 (82)	Malta	0.000 (130)
Algeria	0.151 (39)	Estonia	0.048 (83)	Norway	0.000 (130)
United States of America	0.150 (40)	Thailand	0.038 (84)	Syria	0.000 (130)
Romania	0.146 (41)	Colombia	0.034 (85)	Turkmenistan	0.000 (130)
Bulgaria	0.145 (42)	United Kingdom	0.033 (86)	United Arab Emirates	0.000 (130)
Cote d'Ivoire	0.142 (43)	France	0.031 (89)		
Guatemala	0.137 (44)	Jamaica	0.031 (89)		

Figure 2 shows variability over time in ethnic cleavages for a handful of countries for which we have more than two data points. (Typically, each observation observes the status of a legislature in between elections, so four data points might be separated by three elections.) We can see that ethnic cleavages have increased in New Zealand and Panama, decreased in the Russian Federation, and remained relatively flat in Thailand and Ukraine over the period of observation. Across the entire sample, the mean absolute change in the legislative segmentation index from one period (election) to the next is 0.034, signaling considerable stability.

Box plots in Figure 3 focus on different regions of the world. Here, we find minimal differ-

## Country trajectories of ethnic cleavage over time

Lines show within-country trajectories; points mark observed elections



**Figure 2:** Over time variability in cleavage structures.

ences across medians. However, Asia and especially Africa have considerably higher variability, conforming to standard views in the literature (cited above).

Coverage map outlined in Appendix A (Table ??).

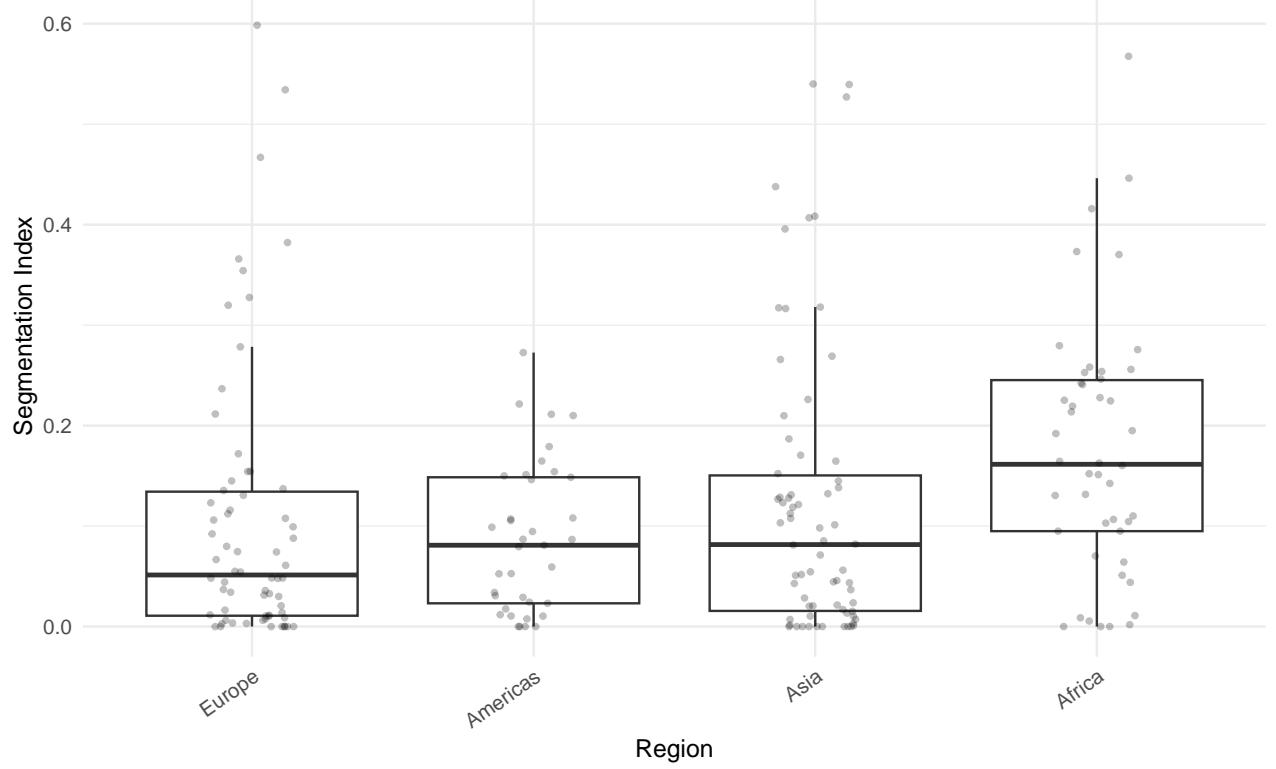
### 3.1 Party and Group Segmentation

Recall that our approach to measurement renders a measure of ethnic segmentation for legislatures (presented above), parties, and groups.

Figure 4 illustrates the distributions of the segmentation index across three units of analysis, revealing a stark contrast: legislative indices are tightly concentrated near zero, indicating minimal ethnic cleavages in most countries, while group- and party-level indices show greater dispersion. This pattern underscores the mechanical interdependence of the indices—where the legislative score is a seat-share-weighted average of party scores—and highlights substantial within-legislature heterogeneity in ethnic alignment. For more details on interdependence between indices, see Appendix XXX.

**How much of the dispersion is within countries?** Because  $S_i^{\text{legislature}}$  aggregates  $S_j^{\text{party}}$  and  $S_i^{\text{group}}$  with country-specific weights (party and group seat shares), variation in the party- and group-level indices reflects two sources: between-country differences in marginals (how fractionalized societies and party systems are) and within-country allocation patterns (how groups and parties line up inside a given legislature). We quantify their relative importance with a weighted variance decomposition. For parties, let  $s_{cjt} = S_j^{\text{party}}$  for party  $j$  in country  $c$  and period  $t$ , weight each observation by its seat share  $w_{cjt} = o_{+j,ct}/n_{ct}$ , and compute the weighted grand mean  $\bar{s} = \sum w_{cjt}s_{cjt}$

### Regional Variability in Ethnic Cleavage



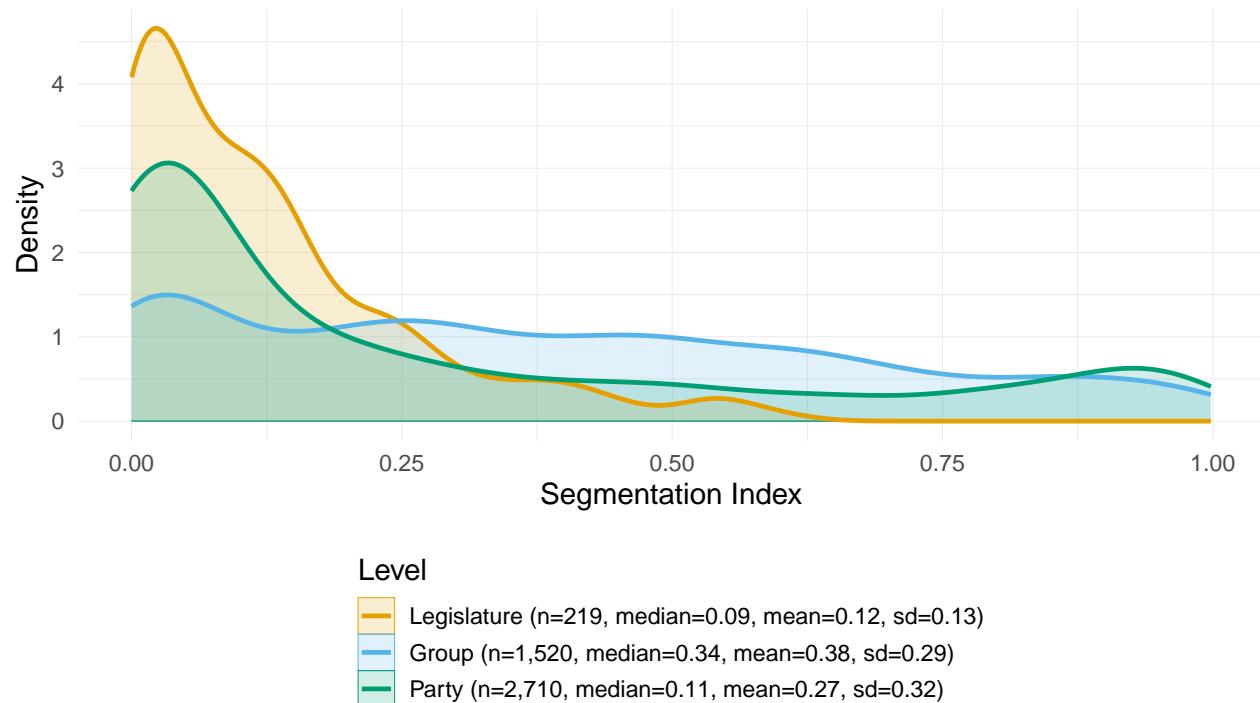
**Figure 3:** Distribution of the segmentation index by region.

and country means  $\bar{s}_c = \sum_{j,t \in c} w_{cjt} s_{cjt} / \sum_{j,t \in c} w_{cjt}$ . The within-country share of variance is

$$\text{WithinShare}_{\text{party}} = \frac{\sum_c \sum_{j,t \in c} w_{cjt} (s_{cjt} - \bar{s}_c)^2}{\sum_c \sum_{j,t \in c} w_{cjt} (s_{cjt} - \bar{s})^2} = 1 - \eta_{\text{between,country}}^2,$$

with the analogous expression for groups using weights  $w_{cit} = o_{i+,ct} / n_{ct}$ . Reporting these as country-weighted  $\eta^2$  (between) and  $1 - \eta^2$  (within) makes the link to the body index transparent, because the weights coincide with those that aggregate components into  $S^{\text{body}}$ . Substantively, we find that a large share of the action is inside polities: the within-country share of the seat-share-weighted variance is 0.562 for party segmentation (between-country: 0.438) and 0.416 for group segmentation (between-country: 0.584). Even where legislative segmentation is low, some parties and groups may be segmented. Likewise, where legislative segmentation is high, some parties and groups may be integrated. These nuances are obscured by the overall legislative segmentation scores.

### Segmentation Index Distributions by Unit of Analysis



**Figure 4:** Density of segmentation index by unit of analysis.

### 3.2 Convergent Validity: Validation Against Existing Measures

In this section, we compare our indices with extant scholarly work. As discussed, most studies judge the ethnic character of politics primarily on the basis of mass political behavior, i.e., on patterns of ethnic voting. Since our measure rests on the ethnic identity of elites, we are comparing two approaches to the same general subject. We expect to find some degree of concordance, but not perfect concordance. Where correlations are weak, this suggests a disjuncture between politics at mass and elite levels (a potential avenue for future research).

To make these comparisons, we draw on studies of ethnic voting and ethnic parties with the broadest coverage, e.g., Houle, Park, and Kenny (2019), Fraenkel and Crawley (2025), Lublin (2014). We may also enlist the V-Party dataset Düpont et al. (2022), where the orientation of individual parties is coded, allowing for a direct comparison with our Ethnic party index. (Lublin’s dataset may also allow for this sort of party-centered comparison.)

**Table 5:** Convergent validity: correlations between our elite-based measures and external indicators. We report Pearson correlations; \*  $p < 0.05$  for Fisher’s z test of non-zero correlation. External measures draw on CPDS ethnic party vote/share (via QoG). Elite-party seat share treats a party as “ethnic” when the modal (non-chamber plurality) ethnic group among its legislators comprises at least =50% of that party.

Measure 1	Measure 2	Pearson R	N
Segmentation Index	Ethnic votes share	0.387*	33
Segmentation Index	Ethnic seats share	0.376*	33
Ethnic-party seat share	Ethnic votes share	0.351*	33
Ethnic-party seat share	Ethnic seats share	0.335	33
Segmentation Index	Ethnic-party seat share	0.734*	130

## 4 Explanations

Having described elite-level ethnic cleavages across the world, we turn from measurement to explanation. Our focus will be at the country level, and hence on the legislative segmentation index.

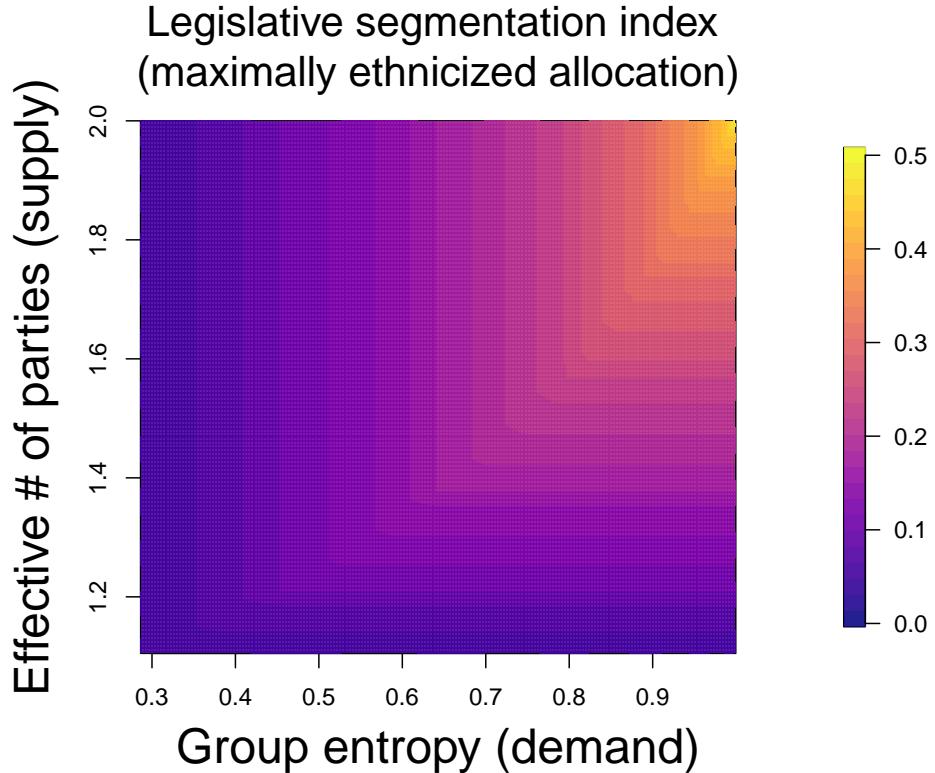
Recall that an ethnic cleavage is defined here as the degree of alignment between ethnic groups and party delegations. This means that there is a mechanical connection to the distribution of ethnic groups and political parties. At the lower limit, where there is only one ethnic group or one political party, there can be no ethnic cleavages. As ethnic groups and political parties multiply, the space available for ethnic cleavages expands. Following previous work, we refer to this as a “compositional” relationship since it depends upon the fit between sociological and institutional configurations (Gerring, Jerzak, and Öncel, 2024).

### 4.1 Simulations

To demonstrate the compositional elements of this puzzle, we explore the full range of counterfactual group and party fractionalization scenarios in Appendix B. Consider the mechanics of sorting: to achieve high segmentation, one needs both distinct items to sort (ethnic diversity) and distinct bins to sort them into (political parties). If either is missing, segmentation collapses.

We formalize this intuition by simulating the maximum possible legislative segmentation index  $S_{body}$  across a grid of “Demand” (group entropy) and “Supply” (effective number of parties). The resulting phase diagram (Figure 5) reveals a stark compositional constraint. High segmentation is

impossible in homogeneous societies (where there is no difference to organize) and equally impossible in single-party systems (where there is no mechanism for sorting). Consequently, the zone of highest ethnic segmentation is structurally confined to the upper-right quadrant: it requires the simultaneous presence of high social entropy and a fragmented party system to capture it.



**Figure 5: The Supply and Demand of Segmentation.** A simulation of the maximum possible legislative segmentation score across varying levels of Group Entropy (Demand) and Effective Number of Parties (Supply). The heatmap shows that high segmentation is a compositional rarity, attainable only when high social diversity interacts with high party fragmentation. By focusing on the maximum possible segmentation, we are, in effect, turning all the knobs to their limits—asking, for each mix of groups and parties, how much segmentation the system could ever sustain—so that later we can read observed scores as using more or less of that built-in potential.

## 4.2 Observational Evidence

To test these relationships with our data, we focus on the dispersion of groups and parties. Dispersion may be operationalized in a variety of ways, e.g., using a Herfindahl index, an entropy index, a largest-share index (the size of the largest group or party), and so forth. All are highly correlated and, unsurprisingly, all produce very similar results. We adopt the Herfindahl index because of its familiarity and ease of interpretation.

In Model 1, Table 6, we test the relationship of the legislative segmentation index to the degree of ethnic and party fractionalization (measured with the Herfindahl index) in each election across the entire sample, clustering standard errors by country. In Model 2, we interact the two variables. Model 3 adds two-way fixed effects.

	Model 1	Model 2	Model 3	Model 4
Party Frac.	0.20 (6.93)*	0.00 (0.21)	-0.05 (-0.66)	-0.01 (-0.29)
Group Frac.	0.38 (13.90)*	-0.01 (-0.14)	-0.16 (-0.90)	-0.03 (-0.48)
Party × Group Frac.		0.61 (5.56)*	0.76 (2.95)*	0.65 (5.61)*
Polyarchy Index				-0.02 (-0.44)
Power Distributed by Social Group				0.00 (0.17)
Access to Public Services by Social Group				-0.01 (-0.71)
Access to State Jobs by Social Group				-0.01 (-0.89)
Access to State Business Opportunities by Social Group				0.01 (1.24)
log(Population)				-0.01 (-1.71)
log(GDP per capita (PPP))				0.01 (1.25)
Country, Percent of Pop. Urbanized				0.00 (-1.40)
Country FE				✓
Year FE				✓
<i>Other statistics</i>				
Countries	130	130	130	122
Observations	219	219	219	204
Adjusted R-squared	0.69	0.76	0.94	0.77

**Table 6:** Outcome: Segmentation Index. Estimator: OLS with clustered standard errors by country. \* indicates  $p < 0.05$ ; t-statistics are in parentheses.

In Model 4, we add a variety of additional predictors, including the Polyarchy index of electoral democracy (Teorell et al., 2019), per capita GDP, PPP, log (World Bank, 2024), urbanization (ibid.), population, log (ibid.), state history (Borcan, Olsson, and Puttermann, 2018), and electoral system rules (Coppedge et al., 2025). We also include V-Dem indicators measuring the distribution of resources — power, services, state jobs, and business opportunities — across social groups (ibid.), understood as a measure of social discrimination.

None of these background factors does a very good job of predicting ethnic cleavages, and the overall model fit is scarcely improved. Moreover, the estimate for the interaction term is stable relative to Model 2, thereby mitigating concerns about omitted-variable bias. This analysis is replicated in a LASSO regression (Figure 14), which yields the same general results: the interaction between group and party fractionalization is consistently selected as non-zero.

We conclude that most of the variability in legislative segmentation scores – roughly three-quarters according to Models 2 and 3 – is attributable to the intersection of demography and party systems, and is therefore compositional, as theorized. To be sure, ethnic and party fractionalization are not entirely independent of each other. The two are very weakly correlated (Pearson’s  $R = 0.17$ ), so there is little risk of collinearity. However, extant work suggests that identities influence party affiliations (Ordeshook and Shvetsova, 1994) and party politics influences identities (Eifert, Miguel, and Posner, 2010). If so, the interaction term may be viewed as an attempt to model that inter-relationship.

## 5 Do Cleavages Matter?

Before concluding, let us return to a question raised briefly at the outset (where citations to the literature may be found). Do ethnic cleavages matter for the quality of governance? The conventional view is that diversity in ascriptive characteristics such as religion, language, and race – bundled together here as *ethnicity* – are problematic for governance. Other studies cast doubt on this pessimistic thesis, leaving the question unresolved. We propose that disparate conclusions in this long-running debate may be reconciled by distinguishing the phenomenon of ethnicity at mass and elite levels.

Extant measures of ethnic politics focus mostly on the characteristics of citizens (as captured by measures of ethnic fractionalization) and voters (as captured by measures of ethnic voting and ethnic parties). Yet, neither of these factors offer a direct indication how politicized ethnic identities might be in a given context. Consider that large ethnic groups are usually geographically segregated. Within each ethnic homeland, candidates for public office – regardless of party – are likely to reflect the ethnic characteristics dominant in that region (Jerzak et al., 2025). This means that voters in a particular district face choices among candidates and parties but not among ethnicities. In this situation, one cannot interpret ethnic voting as an expression of highly politicized – much less polarized – ethnic identities.

However, when latent ethnic cleavages are manifest at elite levels the situation is quite different. Specifically, when parties represented in a national legislature are segmented by ethnicity this is an indication that a demographic feature defines politics at the highest levels. Here, it seems reasonable to regard ethnicity as politicized, perhaps even polarized. Even rich countries with long democratic histories such as Belgium experience the strain of ethnically based political cleavages (De Winter and Baudewyns, 2013).

To probe the thesis that elite-level ethnic cleavages are more problematic than mass-level cleavages we run a series of correlational tests. These should not be mistaken for well-identified causal models, a goal that is beyond reach with the data at hand. Nonetheless, correlational tests offer important clues to causality. And if correlations are predictive rather than causal, this too is important, especially if elite ethnic segmentation offers an early warning of governance failure.

The final point to note about these tests is that they are intended to assess the *relative* importance of various ethnic cleavage measures. If identical tests reveal that one measure is more strongly correlated with governance outcomes than another, this is strong *prima facie* evidence that it is more relevant – causally and/or predictively.

Our empirical strategy begins with three governance outcomes. Democracy is captured by the Polyarchy index (Teorell et al., 2019). Corruption control is captured by the V-Dem index (Coppedge et al., 2025). Political stability is captured by the Political Stability and Absence of Violence/Terrorism index from World Governance Indicators (Kaufmann and Kraay, 2024). Higher scores indicate a more positive outcome – more democracy, less corruption, more stability.

For each outcome, we offer three tests. The first includes extant measures of ethnic politics, introduced in the previous section. The second includes our legislative segmentation index. The third includes all of these measures together. All specifications include two background features that may confound the relationships of interest: per capita GDP (log) and population (log).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Outcome	Democracy	Democracy	Corruption	Corruption	Pol. Stability	Pol. Stability
Segmentation Index		-0.27 (-1.28)		-1.57 (-1.33)		-1.60 (-2.37)*
Group Frac.	0.04 (0.55)	0.14 (1.31)	0.18 (0.46)	0.79 (1.19)	0.03 (0.13)	0.65 (1.74)
Party Frac.	0.48 (6.12)*	0.53 (5.57)*	0.55 (1.15)	0.87 (1.73)	-0.07 (-0.29)	0.26 (0.97)
log(GDP p.c., PPP)	0.10 (4.98)*	0.10 (5.19)*	0.82 (6.97)*	0.83 (7.15)*	0.53 (7.23)*	0.55 (7.66)*
log(Population)	-0.03 (-3.31)*	-0.03 (-3.28)*	-0.19 (-3.28)*	-0.19 (-3.25)*	-0.23 (-6.39)*	-0.23 (-6.15)*
Percent Urban	0.00 (1.00)	0.00 (0.86)	0.01 (1.53)	0.01 (1.44)	0.00 (-0.71)	0.00 (-0.93)
<i>Other statistics</i>						
Countries	122	122	122	122	122	122
Observations	193	193	193	193	195	195
Adjusted R-squared	0.50	0.51	0.55	0.56	0.56	0.57

**Table 7:** Elite-level ethnic cleavage and governance outcome. OLS with heteroskedastic-corrected standard errors. \* indicates  $p < 0.05$ ; t-statistics in parentheses.

## 6 Conclusion

Our analysis advances the study of ethnic politics on three fronts. Conceptually, we recast elite cleavages as a continuous alignment problem and show that an optimal-transport formulation yields a single, bounded, and interpretable scale—the minimal share of seats that would have to be reassigned across parties (holding MPs’ identities fixed) to achieve perfect integration. The same logic provides party-level and group-level diagnostics, allowing researchers to see not only *whether* a system is segmented but also *where* segmentation resides.

Substantively, we find robust evidence for a simple demand–supply account: elite ethnic segregation rises with underlying social heterogeneity and is amplified—rather than created—by party fragmentation; institutional moderators behave in broadly intuitive ways but are less stable than the demand fundamentals. Taken together, these contributions furnish a replicable baseline for cross-national and over-time comparisons, enable targeted evaluation of electoral reforms and power-sharing arrangements, and open a path to cumulative tests linking elite segmentation to downstream outcomes—accountability, clientelism, public-goods provision, and conflict.

Limitations remain—most notably the definitional choices that any coding of ethnicity entails and our focus on represented groups—but the quantitative framework is flexible: transport costs in the cleavage measure can incorporate geography, multi-membership identities can be accommodated, and the design extends naturally to other ascriptive cleavages (religion, language, region) and other arenas (cabinets, local councils).  $\square$

## References

- Abdelal, Rawi et al., eds. (2009): *Measuring Identity: A Guide for Social Scientists*. Cambridge: Cambridge University Press.
- Adida, Claire et al. (2017): “Overcoming or reinforcing coethnic preferences? an experiment on information and ethnic voting”. In: *Quarterly Journal of Political Science*, no. 4, vol. 12, pp. 437–477.
- Alesina, Alberto and Eliana La Ferrara (2005): “Ethnic diversity and economic performance”. In: *Journal of economic literature*, no. 3, vol. 43, pp. 762–800.
- Alford, Robert R (1962): “A suggested index of the association of social class and voting”. In: *Public Opinion Quarterly*, pp. 417–425.
- Arriola, Leonardo R (2012): *Multi-ethnic coalitions in Africa: Business financing of opposition election campaigns*. Cambridge University Press.
- Berman, Bruce, Dickson Eyoh, and Will Kymlicka (2004): *Ethnicity and democracy in Africa*. Ohio University Press.
- Birnir, Jóhanna Kristín (2006): *Ethnicity and electoral politics*. Cambridge University Press.
- Borcan, Oana, Ola Olsson, and Louis Putterman (2018): “State history and economic development: evidence from six millennia”. In: *Journal of Economic Growth*, no. 1, vol. 23, pp. 1–40.
- Chandra, Kanchan (2005): “Ethnic parties and democratic stability”. In: *Perspectives on politics*, no. 2, vol. 3, pp. 235–252.
- (2007): *Why ethnic parties succeed: Patronage and ethnic head counts in India*. Cambridge University Press.
- (2011): “What is an ethnic party?” In: *Party Politics*, no. 2, vol. 17, pp. 151–169.
- Converse, Philip E (1958): “The shifting role of class in political attitudes and behavior”. In: *Readings in social psychology*, vol. 3.
- Coppedge, Michael et al. (2025): *V-dem codebook v15*. Tech. rep. Varieties of Democracy Project.
- Csata, Zsombor, Roman Hlatky, and Amy H Liu (2021): “How to head count ethnic minorities: validity of census surveys versus other identification strategies”. In: *East European Politics*, no. 3, vol. 37, pp. 572–592.
- De Winter, Lieven and Pierre Baudewyns (2013): “Belgium: Towards the breakdown of a nation-state in the heart of Europe?” In: *Pathways from Ethnic Conflict*. Routledge, pp. 20–44.
- Dixit, Avinash and John Londregan (1996): “The determinants of success of special interests in redistributive politics”. In: *the Journal of Politics*, no. 4, vol. 58, pp. 1132–1155.
- Düpont, Nils et al. (2022): “A global perspective on party organizations. Validating the varieties of party identity and organization dataset (V-Party)”. In: *Electoral Studies*, vol. 75, p. 102423.
- Easterly, William and Ross Levine (1997): “Africa’s growth tragedy: policies and ethnic divisions”. In: *The quarterly journal of economics*, pp. 1203–1250.
- Eifert, Benn, Edward Miguel, and Daniel N Posner (2010): “Political competition and ethnic identification in Africa”. In: *American journal of political science*, no. 2, vol. 54, pp. 494–510.
- Elischer, Sebastian (2013): *Political parties in Africa: Ethnicity and party formation*. Cambridge University Press.
- Fraenkel, Jon and Sam Crawley (2025): “Does Ethnic Fractionalization Lead to Polarization? Introducing a new Dataset on Communal Party Affiliations”. In: *Ethnopolitics*, pp. 1–23.
- Geertz, Clifford (1963): “The integrative revolution: primordial sentiments and civil politics in the new states.” In: *Old societies and new states: The quest for modernity in Asia and Africa*, pp. 105–157.

- Gerring, John, Connor T Jerzak, and Erzen Öncel (2024): “The composition of descriptive representation”. In: *American Political Science Review*, no. 2, vol. 118, pp. 784–801.
- Goffman, Erving (1959): *The presentation of self in everyday life*. Doubleday.
- Gould, Lewis (2007): *Grand Old Party: A History of the Republicans*. Random House.
- Hamza, Ahmed and Ko Maeda (2025): “Beyond binary: A new measure of party ethnicization”. In: *Party Politics*, no. 0, vol. 0, p. 13540688251376485. DOI: 10.1177/13540688251376485.
- Horowitz, Donald L (1985): *Ethnic groups in conflict*. Univ of California Press.
- Houle, Christian (2018): “Does ethnic voting harm democracy?” In: *Democratization*, no. 5, vol. 25, pp. 824–842.
- Houle, Christian, Chunho Park, and Paul D Kenny (2019): “The structure of ethnic inequality and ethnic voting”. In: *The Journal of Politics*, no. 1, vol. 81, pp. 187–200.
- Huber, John D (2012): “Measuring Ethnic Voting: Do Proportional Electoral Laws Politicize Ethnicity?” In: *American Journal of Political Science*, no. 4, vol. 56, pp. 986–1001.
- Ishiyama, John (2009): “Do ethnic parties promote minority ethnic conflict?” In: *Nationalism and ethnic politics*, no. 1, vol. 15, pp. 56–83.
- (2012): “Explaining ethnic bloc voting in Africa”. In: *Democratization*, no. 4, vol. 19, pp. 761–788.
- Jerzak, Connor T et al. (2025): “Where Minorities are the Majority: Electoral Rules and Ethnic Representation”. In.
- Kaufmann, Daniel and Aart Kraay (2024): “The worldwide governance indicators”. In.
- Kitschelt, Herbert (2007): *Patrons, Clients and Policies: Patterns of Democratic Accountability and Political Competition*. Cambridge University Press.
- Layman, Geoffrey (2001): *The great divide: Religious and cultural conflict in American party politics*. Columbia University Press.
- Lemarchand, René (1972): “Political clientelism and ethnicity in tropical africa: competing solidarities in nation-building”. In: *American political science review*, no. 1, vol. 66, pp. 68–90.
- Lijphart, Arend (1969): “Consociational democracy”. In: *World politics*, no. 2, vol. 21, pp. 207–225.
- (1977): *Democracy in plural societies: A comparative exploration*. Yale University Press.
- Lipset, Seymour Martin and Stein Rokkan (1967): *Party systems and voter alignments: cross-national perspectives*. International yearbook of political behavior research.
- Liu, Amy H and Jacob I Ricks (2022): *Ethnicity and politics in Southeast Asia*. Cambridge University Press.
- Lott, John and Cédric Villani (2009): “Ricci curvature for metric-measure spaces via optimal transport”. In: *Annals of Mathematics*, pp. 903–991.
- Lublin, David (2014): *Minority Rules: Electoral Systems, Decentralization, and Ethnoregional Party Success*. Oxford University Press.
- Madrid, Raúl L (2012): *The rise of ethnic politics in Latin America*. Cambridge University Press.
- Marks, Gary et al. (2023): “The social bases of political parties: A new measure and survey”. In: *British Journal of Political Science*, no. 1, vol. 53, pp. 249–260.
- Marquardt, Kyle L and Yoshiko M Herrera (2015): “Ethnicity as a variable: an assessment of measures and data sets of ethnicity and related identities”. In: *Social Science Quarterly*, no. 3, vol. 96, pp. 689–716.
- Martínez Mauri, Mònica (2012): “The social and political construction of racial and ethnic categories in National censuses of Panama, 1911-2010”. In: *Everlasting Countdowns: Race, Ethnicity and National Censuses in Latin American States*, pp. 155–184.

- Ordeshook, Peter C and Olga V Shvetsova (1994): "Ethnic heterogeneity, district magnitude, and the number of parties". In: *American journal of political science*, pp. 100–123.
- Page, Scott (2008): *The difference: How the power of diversity creates better groups, firms, schools, and societies-new edition*. Princeton University Press.
- Posner, Daniel N (2004): "Measuring ethnic fractionalization in Africa". In: *American journal of political science*, no. 4, vol. 48, pp. 849–863.
- Rabushka, Alvin and Kenneth A Shepsle (1972): *Politics in plural societies*. Charles E. Merrill.
- Rae, Douglas W. and Michael Taylor (1970): *The analysis of political cleavages*. Yale University Press.
- Reilly, Benjamin (2001): *Democracy in Divided Societies: Electoral Engineering for Conflict Management*. Cambridge University Press.
- (2021): "Cross-ethnic voting: An index of centripetal electoral systems". In: *Government and Opposition*, no. 3, vol. 56, pp. 465–484.
- Rustow, Dankwart A (1970): "Transitions to democracy: Toward a dynamic model". In: *Comparative politics*, no. 3, vol. 2, pp. 338–63.
- Rydgren, Jens (2007): "The sociology of the radical right". In: *Annu. Rev. Sociol.* No. 1, vol. 33, pp. 241–262.
- Schrover, Marlou (2010): "Pillarization, multiculturalism and cultural freezing. Dutch migration history and the enforcement of essentialist ideas". In: *BMGN-Low Countries Historical Review*, no. 2-3, vol. 125, pp. 329–354.
- Silbey, Joel H and Allan G Bogue (1978): *The history of American electoral behavior*. Princeton University Press.
- Strijbis, Oliver and Michal Kotnarowski (2015): "Measuring the electoral mobilization of ethnic parties: Towards comparable indicators". In: *Party Politics*, no. 3, vol. 21, pp. 456–469.
- Sturm, Johan et al. (1998): "Educational Pluralism-a historical study of so-called'pillarization'in the Netherlands, including a comparison with some developments in South African education". In: *Comparative education*, no. 3, vol. 34, pp. 281–297.
- Teorell, Jan et al. (2019): "Measuring polyarchy across the globe, 1900–2017". In: *Studies in Comparative International Development*, vol. 54, pp. 71–95.
- Van Cott, Donna Lee (2007): *From movements to parties in Latin America: The evolution of ethnic politics*. Cambridge University Press.
- World Bank (2024): *World development indicators 2024*. International Economics Department. Development Data Group and World Bank.

## 7 Appendix A

### 7.1 Additional Discussion

**Mechanical link between the three indices.** Let  $O = [o_{ij}]$  be the groups  $\times$  parties seat matrix in a chamber with  $n = \sum_{i,j} o_{ij}$  MPs, row totals  $o_{i+}$  (groups) and column totals  $o_{+j}$  (parties). The “perfect-integration” target preserves these marginals:  $R = [r_{ij}]$  with  $r_{ij} = o_{i+}o_{+j}/n$ . The body-level segmentation index is the (normalized) total-variation distance under uniform 0-1 party switching costs:

$$S^{\text{body}} = \frac{1}{2n} \sum_{i,j} |o_{ij} - r_{ij}|.$$

Define a party’s segmentation score and a group’s segmentation score by the same distance, locally normalized:

$$S_j^{\text{party}} = \frac{1}{2o_{+j}} \sum_i |o_{ij} - r_{ij}|, \quad S_i^{\text{group}} = \frac{1}{2o_{i+}} \sum_j |o_{ij} - r_{ij}|.$$

By construction,

$$S^{\text{body}} = \sum_j \frac{o_{+j}}{n} S_j^{\text{party}} = \sum_i \frac{o_{i+}}{n} S_i^{\text{group}}.$$

Hence, the chamber score is simultaneously a seat-share-weighted average of party scores and a group-share-weighted average of group scores. Two implications follow. First,  $S^{\text{body}}$  is a convex combination of the components and therefore lies between their minimum and maximum; body-level dispersion is mechanically tighter than dispersion across parties or groups. Second, the same chamber score can arise from distinct micro-configurations—e.g., uniformly moderate party (or group) scores versus a mixture of very high and very low scores that average out—which is exactly the heterogeneity the party- and group-level indices reveal.

### 7.2 Additional Analyses

**Table 8:** Cases (legislature-years) flagged as below the coverage threshold (0.50) prior to dropping. Countries shown only if at least one case is below threshold.

Note: Percent missing is computed among country-years with finite coverage.

Country	Cases below threshold	Units with coverage	% missing
Nigeria	2	2	96.8
Timor-Leste (East Timor)	2	2	93.8
Liberia	2	2	93.2
Cyprus	2	2	91.4
Korea, North	2	2	90.5
Kyrgyzstan	2	2	81.7
Bahrain	2	2	78.8
Luxembourg	1	1	98.3
Suriname	1	1	98.0
Congo	1	1	97.0
Netherlands	1	2	69.3
Senegal	1	2	65.7
Namibia	1	2	60.3
Morocco	1	2	59.2
Gabon	1	2	58.1

**Table 8:** Cases (legislature-years) flagged as below the coverage threshold (0.50) prior to dropping. Countries shown only if at least one case is below threshold.

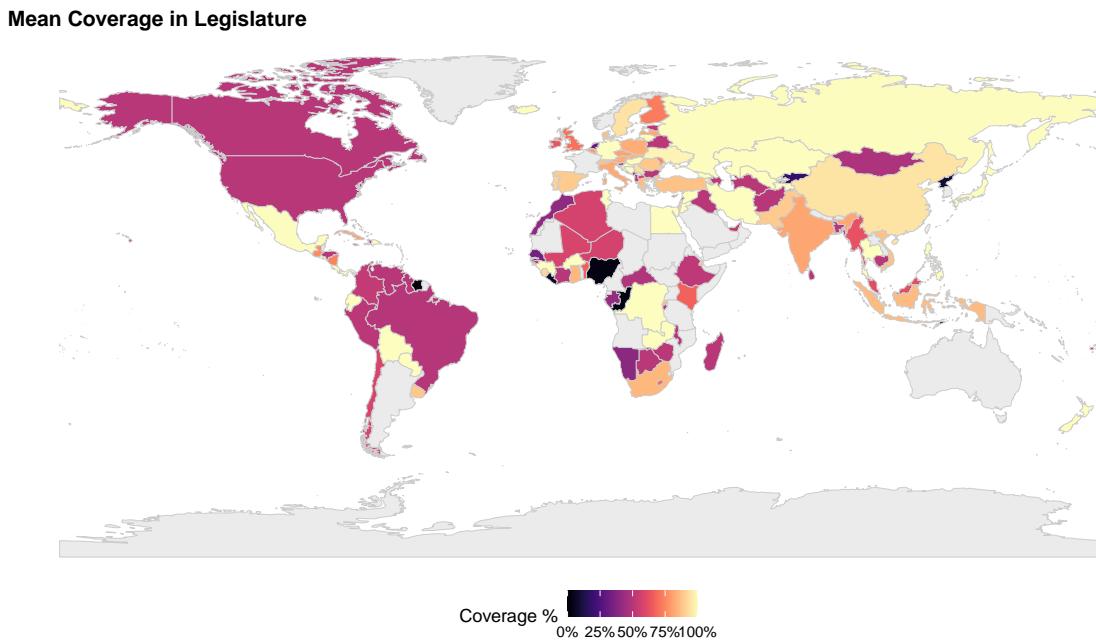
Note: Percent missing is computed among country-years with finite coverage. (continued)

Country	Cases below threshold	Units with coverage	% missing
Montenegro	1	2	53.6
Mongolia	1	2	52.0
Central African Republic	1	2	51.4
Malawi	1	2	50.4
Brazil	1	2	49.9
Cote d'Ivoire (Ivory Coast)	1	2	49.8
Venezuela, Bolivarian Republic of	1	2	49.7
Bangladesh	1	2	49.7
United States of America	1	2	49.7
Lebanon	1	2	49.6
Bulgaria	1	2	49.6
Burundi	1	2	49.6
Slovenia	1	2	49.4
Iraq	1	2	49.4
Canada	1	2	49.4
Madagascar	1	2	49.3
Sri Lanka	1	2	49.3
Guyana	1	2	49.2
Honduras	1	2	49.2
Cambodia	1	2	49.2
Turkmenistan	1	2	49.2
Botswana	1	2	49.1
Belarus	1	2	49.1
Zimbabwe	1	2	49.1
Albania	1	2	48.9
Peru	1	2	48.8
Trinidad and Tobago	1	2	48.8
Cape Verde	1	2	48.6
Guinea-Bissau	1	2	48.5
Estonia	1	2	48.5
Haiti	1	2	48.4
Azerbaijan	1	2	48.4
Ethiopia	1	2	48.3
Fiji	1	2	48.0
United Arab Emirates	1	2	47.5
Colombia	1	2	47.0
Niger	1	2	42.8
Algeria	1	2	42.6
Chile	1	2	41.7
Macedonia	1	2	41.5
Malaysia	1	2	39.0
Myanmar	1	2	37.4
Ireland	1	2	36.7
Jamaica	1	2	34.1
Kenya	1	2	32.9
United Kingdom (Great Britain)	1	2	29.7
Finland	1	2	27.2
Moldova, Republic of	1	3	25.7

**Table 8:** Cases (legislature-years) flagged as below the coverage threshold (0.50) prior to dropping. Countries shown only if at least one case is below threshold.

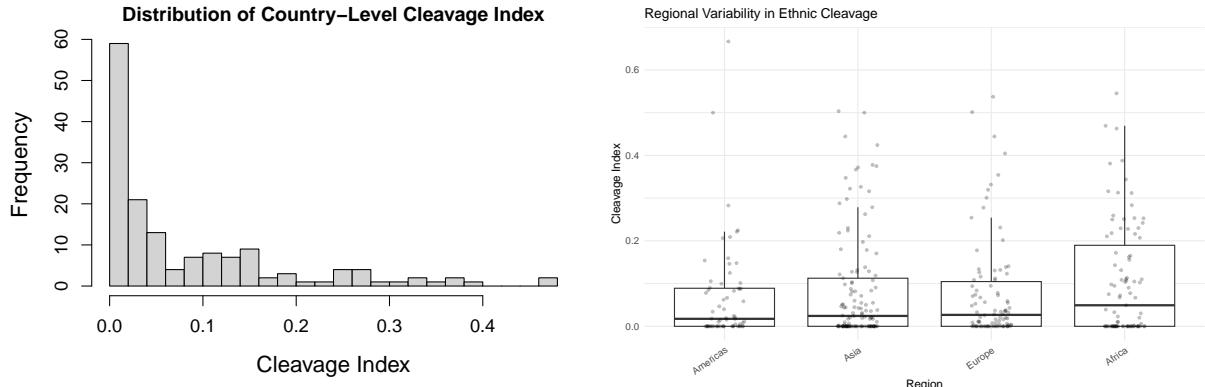
Note: Percent missing is computed among country-years with finite coverage. (continued)

Country	Cases below threshold	Units with coverage	% missing
Latvia		1	3 18.0

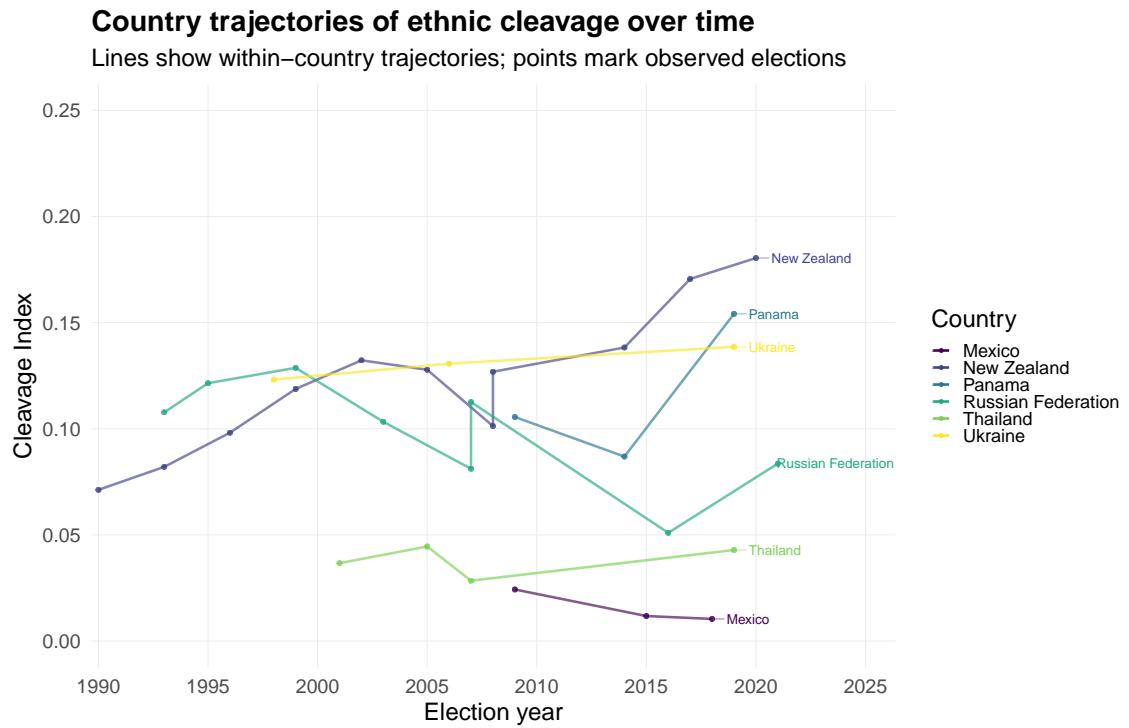


**Figure 6:** LEFT. Coverage, by country.

### 7.3 Results, Expert Only



**Figure 7:** LEFT. Distribution of the segmentation index across country-elections in the sample. The index is bounded between 0 (perfect integration) and 1 (perfect segregation); the distribution is heavily skewed to the right, indicating that most national legislatures in the dataset exhibit low levels of ethnic segregation among their party delegations. RIGHT. Distribution by region.

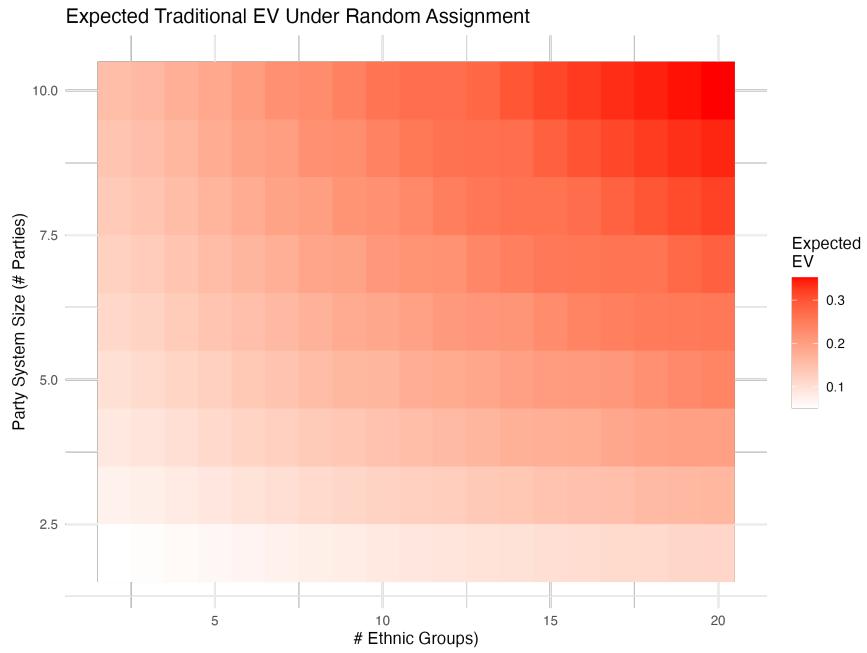


**Figure 8:** Over time variability in cleavage structures.

	Model 1	Model 2	Model 3	Model 4
Party Fractionalization	0.15 (8.24)*	0.01 (0.98)	0.01 (0.61)	0.02 (1.60)
Group Fractionalization	0.27 (12.17)*	0.01 (0.70)	0.04 (0.86)	0.00 (0.33)
Party Times Group Fractionalization		0.63 (13.16)*	0.56 (6.40)*	0.63 (12.72)*
Polyarchy Index				0.02 (0.98)
Power Distributed by Social Group				-0.01 (-1.35)
Access to Public Services by Social Group				0.00 (-0.53)
Access to State Jobs by Social Group				0.00 (-0.12)
Access to State Business Opportunities by Social Group				0.00 (0.11)
log(Population)				0.00 (-1.27)
log(GDP per capita (PPP))				0.00 (-0.75)
Country, Percent of Pop. Urbanized				0.00 (0.63)
Country FE				✓
Year FE				✓
<i>Other statistics</i>				
Countries	153	153	153	139
Observations	374	374	374	337
Adjusted R-squared	0.57	0.75	0.80	0.75

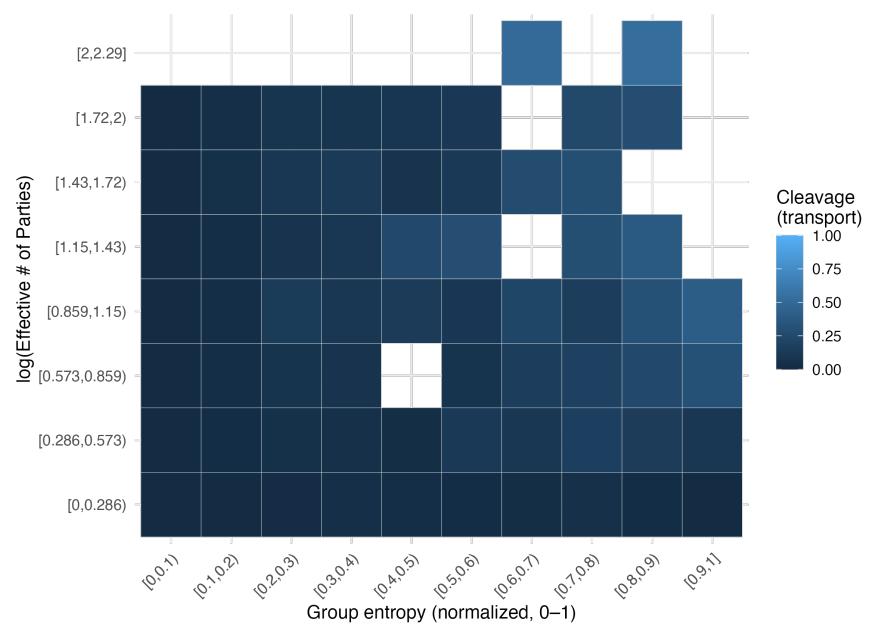
**Table 9:** Outcome: Cleavage Index. Estimator: OLS with clustered standard errors by country. \* indicates  $p < 0.05$ ; t-statistics are in parentheses.

## 7.4 Appendix B: Simulations



**Figure 9:** Heatmap.

While cleavage dynamics appear stable, compositional features of the country population seem to be more important drivers, as seen in Figure 10. Here, we observe a pronounced upper-right gradient: the segmentation index increases monotonically with both normalized group entropy and the number of effective parties; high segregation appears almost exclusively where social diversity (demand) and party-system fragmentation (supply) are simultaneously large, in line with our demand–supply account.



**Figure 10:** Heat map: Group entropy and number of effective parties.

## 8 Appendix C

### —Groups—

*Afghanistan (2 elections, 774 people total): Pashtun (343, 46.7*

*Albania (2 elections, 317 people total): Albanian (196, 99.5*

*Algeria (2 elections, 950 people total): Arab (357, 67.4*

*Armenia (2 elections, 321 people total): Armenian (304, 98.4*

*Austria (2 elections, 470 people total): Austrian (463, 98.7*

*Azerbaijan (2 elections, 168 people total): Azeri (139, 97.9*

*Bahrain (2 elections, 150 people total): Sunni Arab (13, 76.5*

*Bangladesh (2 elections, 490 people total): Bengali (404, 99.5*

*Belarus (2 elections, 300 people total): Belarusian (151, 88.3*

*Belgium (2 elections, 337 people total): Flemish (188, 56.8*

*Benin (2 elections, 249 people total): Aja (45, 40.9*

*Bolivia (2 elections, 260 people total): Mestizo (151, 58.1*

*Bosnia and Herzegovina (2 elections, 191 people total): Bosniak (78, 42.9*

*Botswana (2 elections, 124 people total): Tswana (58, 74.4*

*Brazil (2 elections, 1173 people total): Pardos (Multiracial) (770, 94.8*

*Bulgaria (2 elections, 567 people total): Bulgarian (296, 91.1*

*Burkina Faso (2 elections, 345 people total): Mossi (133, 49.8*

*Burundi (2 elections, 301 people total): Hutu (99, 64.7*

*Cambodia (2 elections, 201 people total): Khmer (194, 97.0*

*Canada (2 elections, 675 people total): Canadian (400, 87.5*

*Cape Verde (2 elections, 149 people total): Mestizo/Mulatto (147, 98.7*

*Central African Republic (2 elections, 149 people total): Gbaya (42, 47.2*

*Chile (2 elections, 334 people total): White (233, 98.3*

*China, People's Republic of (2 elections, 5659 people total): Han Chinese (4440, 89.5*

*Colombia (2 elections, 575 people total): Mestizo (352, 93.6*

*Congo (2 elections, 292 people total): Mbochi (3, 75.0*

*Congo, Democratic Republic of the (2 elections, 554 people total): Anamongo (56, 10.6*

*Costa Rica (2 elections, 196 people total): White or Mestizo (193, 99.0*

*Cote d'Ivoire (Ivory Coast) (2 elections, 537 people total): Akan (109, 34.0*

*Croatia (2 elections, 385 people total): Croat (349, 97.2*

*Cuba (2 elections, 1284 people total): White (550, 43.1*

*Cyprus (2 elections, 125 people total): Greek Cypriot (92, 100.0*

*Czech Republic (2 elections, 456 people total): Czech (322, 76.3*

*Denmark (2 elections, 418 people total): Danish (395, 97.8*

*Djibouti (2 elections, 187 people total): Somali (82, 56.9*

*Dominican Republic (2 elections, 445 people total): Mulatto (182, 44.2*

*Ecuador (2 elections, 399 people total): Mestizo (155, 50.2*

*Egypt (2 elections, 1491 people total): Egyptian (1453, 99.4*

*El Salvador (2 elections, 248 people total): Mestizo (190, 81.5*

*Estonia (2 elections, 242 people total): Estonian (151, 92.1*

*Ethiopia (2 elections, 1203 people total): Oromo (321, 47.6*

*Fiji (2 elections, 68 people total): iTaukei (36, 57.1*

*Finland (2 elections, 437 people total): Finnish (386, 93.2*

*France (2 elections, 1273 people total): French (1196, 94.1*

*Gabon (2 elections, 163 people total): Fang (37, 36.6*

*Gambia (2 elections, 139 people total): Bambara (40, 28.8*

*Georgia (2 elections, 300 people total): Georgian (287, 95.7*

*Germany (2 elections, 1423 people total): German (1387, 97.5*

*Ghana (2 elections, 587 people total): Akan (333, 58.1*

*Greece (2 elections, 627 people total): Greek (585, 99.3*

*Guatemala (2 elections, 431 people total): Mestizo (205, 84.7*

*Guinea (2 elections, 208 people total): Fulani(Peul) (70, 33.7*

*Guinea-Bissau (2 elections, 133 people total): Fulani (33, 26.2*

*Guyana (2 elections, 158 people total): Black (38, 46.3*

*Haiti (2 elections, 317 people total): Black (190, 99.5*

*Honduras (2 elections, 312 people total): Mestizo (199, 97.1*

*Hungary (2 elections, 398 people total): Hungarian (391, 98.2*

*Iceland (2 elections, 163 people total): White (163, 100.0*

*India (2 elections, 1274 people total): Hindi (Indo-Aryan, Central) (404, 33.1*

*Indonesia (2 elections, 1238 people total): Javanese (437, 36.0*

*Iran, Islamic Republic of (2 elections, 732 people total): Persian (384, 57.6*

*Iraq (2 elections, 403 people total): Arab (288, 83.7*

*Ireland (2 elections, 342 people total): Irish (316, 99.1*

*Israel (2 elections, 248 people total): Ashkenazi (144, 59.8*

*Italy (2 elections, 1030 people total): Italian (1026, 99.6*

*Jamaica (2 elections, 191 people total): Black (148, 88.6*

*Japan (2 elections, 955 people total): Japanese (955, 100.0*

*Jordan (2 elections, 333 people total): Arab (317, 97.8*

*Kazakhstan (2 elections, 362 people total): Kazakh (273, 76.7*

*Kenya (2 elections, 969 people total): Kikuyu (170, 18.2*

*Korea, North (2 elections, 186 people total): Korean (184, 100.0*

*Kosovo (2 elections, 300 people total): Albanian (255, 92.4*

*Kuwait (1 elections, 66 people total): Kuwaiti (66, 100.0*

*Kyrgyzstan (2 elections, 307 people total): Kyrgyz (198, 96.1*

*Latvia (3 elections, 527 people total): Latvian (335, 73.8*

*Lebanon (2 elections, 325 people total): Maronite (138, 49.5*

*Lesotho (2 elections, 262 people total): Sotho (262, 100.0*

*Liberia (2 elections, 256 people total): Kpelleh (68, 37.6*

*Libya (Libyan Arab Jamahiriya) (2 elections, 275 people total): Arab (274, 100.0*

*Lithuania (2 elections, 282 people total): Lithuanian (268, 95.0*

*Luxembourg (2 elections, 186 people total): Luxemburger (121, 97.6*

*Macedonia (2 elections, 309 people total): Macedonian (223, 74.1*

*Madagascar (2 elections, 327 people total): Malagasy (289, 91.2*

*Malawi (2 elections, 551 people total): Chewa (280, 66.7*

*Malaysia (2 elections, 441 people total): Malay (211, 51.2*

*Mali (2 elections, 381 people total): Bambara (112, 56.6*

*Malta (2 elections, 165 people total): Maltese (151, 93.2*

*Mauritania (2 elections, 99 people total): Haratine (Black Moor, Arab-Berber-Negroid) (2, 50.0*

*Mauritius (2 elections, 169 people total): Hindu (79, 52.3*

*Mexico (3 elections, 1499 people total): White/Mestizo (1462, 97.5*

*Moldova, Republic of (3 elections, 355 people total): Moldovan (225, 67.6*

*Mongolia (2 elections, 83 people total): Khalkha (61, 74.4*

*Montenegro (2 elections, 223 people total): Montenegrin (121, 74.7*

*Morocco (2 elections, 811 people total): Arab (702, 95.0*

*Myanmar (2 elections, 1093 people total): Burman (525, 73.5*

*Namibia (2 elections, 198 people total): Ovambo (82, 54.7*

*Netherlands (2 elections, 340 people total): Dutch (203, 91.9*

*New Zealand (11 elections, 1280 people total): European (991, 77.4*

*Nicaragua (2 elections, 251 people total): Mestizo (226, 92.2*

*Niger (2 elections, 377 people total): Haoussa (101, 58.0*

*Nigeria (2 elections, 597 people total): Hausa (178, 43.3*

*Norway (2 elections, 383 people total): Norwegian (377, 100.0*

*Oman (2 elections, 255 people total): Omani Arab (208, 100.0*

*Pakistan (2 elections, 701 people total): Punjabi (359, 52.5*

*Palestinian Territory, Occupied (2 elections, 264 people total): Palestinian (264, 100.0*

*Panama (3 elections, 214 people total): Mestizo (155, 72.4*

*Paraguay (2 elections, 286 people total): Mestizo (167, 58.4*

*Peru (2 elections, 320 people total): Mestizo (137, 74.9*

*Philippines (2 elections, 685 people total): Malay (624, 93.1*

*Poland (2 elections, 986 people total): Polish (973, 99.4*

*Portugal (3 elections, 822 people total): Portuguese (786, 99.9*

*Puerto Rico (2 elections, 161 people total): White Hispanic (142, 92.2*

*Qatar (2 elections, 114 people total): Qatari Arab (114, 100.0*

*Romania (2 elections, 703 people total): Romanian (602, 90.7*

*Russian Federation (8 elections, 3591 people total): Russian (2881, 80.2*

*Rwanda (2 elections, 289 people total): Hutu (105, 51.0*

*Saudi Arabia (2 elections, 396 people total): Arab (264, 100.0*

*Senegal (2 elections, 198 people total): Wolof (73, 70.9*

*Serbia (2 elections, 607 people total): Serbian (551, 94.2*

*Sierra Leone (2 elections, 438 people total): Mende (89, 23.4*

*Singapore (2 elections, 231 people total): Chinese (172, 78.5*

*Slovakia (2 elections, 349 people total): Slovak (315, 91.8*

*Slovenia (2 elections, 235 people total): Slovenian (144, 94.7*

*Solomon Islands* (2 elections, 125 people total): Malaita (40, 35.4

*Somalia* (2 elections, 288 people total): Somali (279, 100.0

*Somaliland* (1 elections, 113 people total): Somali (113, 100.0

*South Africa* (2 elections, 847 people total): Black African (577, 71.9

*South Korea* (2 elections, 627 people total): Korean (627, 100.0

*Spain* (2 elections, 831 people total): Castellanos (582, 83.3

*Sri Lanka* (2 elections, 456 people total): Sinhalese (283, 80.4

*Suriname* (1 elections, 71 people total): Mixed (1, 100.0

*Sweden* (2 elections, 765 people total): Swedish (678, 91.3

*Switzerland* (2 elections, 523 people total): Swiss Germans (360, 69.2

*Syria (Syrian Arab Republic)* (2 elections, 405 people total): Arab (391, 100.0

*Taiwan* (2 elections, 225 people total): Chinese (213, 94.7

*Tajikistan* (2 elections, 184 people total): Tajik (180, 98.9

*Thailand* (4 elections, 2000 people total): Thai (1936, 96.8

*Timor-Leste (East Timor)* (2 elections, 233 people total): Malay (Polynesian) (4, 50.0

*Trinidad and Tobago* (2 elections, 108 people total): East Indian (30, 43.5

*Tunisia (2 elections, 486 people total): Arab (481, 99.8*

*Turkey (2 elections, 1131 people total): Turkish (838, 80.3*

*Turkmenistan (2 elections, 392 people total): Turkmen (354, 91.7*

*Ukraine (3 elections, 1349 people total): Ukrainian (1030, 78.0*

*United Arab Emirates (2 elections, 77 people total): Emirati Arab (77, 100.0*

*United Kingdom (Great Britain) (2 elections, 1283 people total): White (1203, 94.2*

*United States of America (2 elections, 981 people total): White (599, 81.9*

*Uruguay (2 elections, 274 people total): White (272, 99.6*

*Uzbekistan (2 elections, 352 people total): Uzbek (315, 91.8*

*Venezuela, Bolivarian Republic of (2 elections, 394 people total): Mestizo (265, 97.1*

*Viet Nam (2 elections, 971 people total): Kinh (788, 84.6*

*Yemen (2 elections, 688 people total): Arab (648, 97.9*

*Zambia (2 elections, 372 people total): Bemba (151, 44.7*

*Zimbabwe (2 elections, 477 people total): Shona (221, 75.4*

—Parties—

*Afghanistan (2 elections, 774 people total): Jamiat-e Islami-ye (212, 27.5*

*Albania (2 elections, 317 people total): Socialist Party of Albania (150, 47.3*

*Algeria (2 elections, 950 people total): Independents (312, 32.9*

*Argentina (2 elections, 336 people total): Justicialist Party - Front for Victory (134, 40.0*

*Armenia (2 elections, 321 people total): Republican Party of Armenia (197, 61.6*

*Austria (2 elections, 470 people total): Austrian People's Party (212, 45.9*

*Azerbaijan (2 elections, 168 people total): New Azerbaijan Party (137, 82.0*

*Bahrain (2 elections, 150 people total): The Gathering of National Unity (42, 28.8*

*Bangladesh (2 elections, 490 people total): Bangladesh Awami League (387, 79.0*

*Belarus (2 elections, 300 people total): Block Alexander Lukashenko (280, 93.3*

*Belgium (2 elections, 337 people total): Socialist Party (66, 19.8*

*Benin (2 elections, 249 people total): Cauri Forces for an Emerging Benin (108, 43.5*

*Bolivia (2 elections, 260 people total): Movimiento Al Socialismo - MAS (88, 33.8*

*Bosnia and Herzegovina (2 elections, 191 people total): Party for Democratic Action (40, 20.9*

*Botswana (2 elections, 124 people total): Botswana Democratic Party (88, 71.5*

*Brazil (2 elections, 1173 people total): Brazilian Democratic Movement Party (198, 16.9*

*Bulgaria (2 elections, 567 people total):* Citizens for European Development of Bulgaria (GERB) (273, 48.5

*Burkina Faso (2 elections, 345 people total):* Congress for Democracy and Progress (CDP) (174, 51.0

*Burundi (2 elections, 301 people total):* National Council for the Defense of Democracy (CNDD-FDD) (111, 37.2

*Cambodia (2 elections, 201 people total):* Cambodia People Party (154, 76.6

*Canada (2 elections, 675 people total):* The Conservative Party (270, 40.1

*Cape Verde (2 elections, 149 people total):* African Party for the Independence of Cape Verde (88, 59.1

*Central African Republic (2 elections, 149 people total):* National Convergence 'Kwa Na Kwa' (KNK) (111, 74.5

*Chile (2 elections, 334 people total):* National Renewal (85, 25.8

*China, People's Republic of (2 elections, 5659 people total):* Communist Party of China (4774, 92.4

*Colombia (2 elections, 575 people total):* Social Party of National Unity (154, 26.9

*Congo (2 elections, 292 people total):* Congolese Workers Party (171, 58.8

*Congo, Democratic Republic of the (2 elections, 554 people total):* People's Party for Reconstruction and Democracy (174, 31.4

*Costa Rica (2 elections, 196 people total):* National Liberation Party (91, 47.4

*Côte d'Ivoire (Ivory Coast) (2 elections, 537 people total): Rally of Houphouetists for Democracy and Peace (RHDP) (190, 35.4*

*Croatia (2 elections, 385 people total): Croatian Democratic Union (178, 46.2*

*Cuba (2 elections, 1284 people total): Cuban Communist Party (PCC) (1241, 100.0*

*Cyprus (2 elections, 125 people total): Progressive Party for the Working People (A.K.E.L.) (10, 45.5*

*Czech Republic (2 elections, 456 people total): Czech Social Democratic Party - ČSSD (98, 21.8*

*Denmark (2 elections, 418 people total): Social Democratic Party (Socialdemokraterne) (102, 24.8*

*Djibouti (2 elections, 187 people total): People's Rally for Progress (RPP) (149, 80.5*

*Dominican Republic (2 elections, 445 people total): Dominican Liberation Party (PLD) (251, 56.8*

*Ecuador (2 elections, 399 people total): Alianza País (193, 49.1*

*Egypt (2 elections, 1491 people total): The National Democratic Party (355, 24.0*

*El Salvador (2 elections, 248 people total): Farabundo Martí National Liberation Front (FMLN) (92, 37.1*

*Estonia (2 elections, 242 people total): Estonian Reform Party (91, 37.8*

*Ethiopia (2 elections, 1203 people total): Ethiopian People's Revolutionary Democratic Front (982, 81.6*

*Fiji (2 elections, 68 people total): Soqosoqo Duavata ni Lewenivanua (67, 98.5*

*Finland (2 elections, 437 people total): National coalition party (98, 22.4*

*France (2 elections, 1273 people total): The Republicans (548, 43.3*

*Gabon (2 elections, 163 people total): Gabonese Democratic Party (PDG) (151, 92.6*

*Gambia (2 elections, 139 people total): United Democratic Party (104, 74.8*

*Georgia (2 elections, 300 people total): Georgian Dream-Democratic Georgia-Democratic Georgia (114, 38.0*

*Germany (2 elections, 1423 people total): Christian Democratic Union (CDU)/ Christian Social Union (CSU) (501, 35.3*

*Ghana (2 elections, 587 people total): New Patriotic Party (325, 55.8*

*Greece (2 elections, 627 people total): New Democracy (235, 38.1*

*Guatemala (2 elections, 431 people total): National Unity of Hope (153, 36.1*

*Guinea (2 elections, 208 people total): Union of Democratic Forces of Guinea (UFDG) (83, 40.3*

*Guinea-Bissau (2 elections, 133 people total): African Party for the Independence of Guinea and Cape Verde (98, 74.2*

*Guyana (2 elections, 158 people total): People's Progressive Party (87, 55.1*

*Haiti (2 elections, 317 people total): Patriotic Unity Party (125, 40.7*

*Honduras (2 elections, 312 people total): National Party (172, 57.7*

*Hungary (2 elections, 398 people total): Fidesz-Christian Democratic People's Party (268, 67.3*

*Iceland (2 elections, 163 people total): Independence Party (68, 41.7*

*India (2 elections, 1274 people total): Bharatiya Janata Party (BJP) (388, 30.6*

*Indonesia (2 elections, 1238 people total): Indonesian Democratic Party–Struggle (265, 21.4*

*Iran, Islamic Republic of (2 elections, 732 people total): Islamic Coalition Party (265, 36.6*

*Iraq (2 elections, 403 people total): Iraqi National Alliance (293, 73.6*

*Ireland (2 elections, 342 people total): Fine Gael (130, 38.0*

*Israel (2 elections, 248 people total): Likud (79, 32.0*

*Italy (2 elections, 1030 people total): Five Star Movement (277, 26.9*

*Jamaica (2 elections, 191 people total): Peoples National Party (104, 55.9*

*Japan (2 elections, 955 people total): Liberal Democratic Party (434, 45.4*

*Jordan (2 elections, 333 people total): Independent (14, 24.1*

*Kazakhstan (2 elections, 362 people total): Nur-Otan (347, 96.9*

*Kenya (2 elections, 969 people total): Orange Democratic Movement (ODM) (238, 54.2*

*Korea, North (2 elections, 186 people total): Worker's Party of Korea (125, 96.2*

*Kosovo (2 elections, 300 people total): Democratic Party of Kosovo (DPK) (55, 31.2*

*Kyrgyzstan (2 elections, 307 people total): Social Democratic Party of Kyrgyzstan (SDPK) (31, 23.5*

*Latvia (3 elections, 527 people total): Alliance of Political Organizations (Harmony Center) (143, 33.3*

*Lebanon (2 elections, 325 people total): Future Movement (54, 22.9*

*Lesotho (2 elections, 262 people total): Democratic Congress (69, 41.1*

*Liberia (2 elections, 256 people total): Congress for Democratic Change (4, 28.6*

*Lithuania (2 elections, 282 people total): Homeland Union - Lithuanian Christian Democrats (81, 28.7*

*Luxembourg (2 elections, 186 people total): Christian Social Peoples' Party (57, 48.7*

*Macedonia (2 elections, 309 people total): Internal- Macedonian Revolutionary Organization- Democratic Party for macedonian national unity (IMRO- DPMNU) (76, 52.1*

*Madagascar (2 elections, 327 people total): Determined Malagasy Youth - TGV (53, 24.3*

*Malawi (2 elections, 551 people total): Democratic Progressive Party (DPP) (130, 68.4*

*Malaysia (2 elections, 441 people total): United Malays National Organisation (UMNO) (94, 32.4*

*Mali (2 elections, 381 people total): Alliance for Democracy in Mali-Pan-African Party for Liberty, Solidarity and Justice (ADEMA-PASJ) (65, 38.7*

*Malta (2 elections, 165 people total): Labour Party (51, 50.5*

*Mauritania (2 elections, 99 people total): Union for the Republic (39, 78.0*

*Mauritius (2 elections, 169 people total): Mauritian Labour Party (MLP) (35, 35.7*

*Mexico (3 elections, 1499 people total): Partido Revolucionario Institucional - PRI (462, 30.8*

*Moldova, Republic of (3 elections, 355 people total): Communist Party of the Republic of Moldova (102, 41.5*

*Mongolia (2 elections, 83 people total): Mongolian People's Revolutionary Party (47, 63.5*

*Montenegro (2 elections, 223 people total): Democratic Party of Socialists (64, 48.9*

*Morocco (2 elections, 811 people total): Justice and Development Party (PJD) (145, 38.0*

*Myanmar (2 elections, 1093 people total): Union Solidarity and Development Party (390, 76.9*

*Namibia (2 elections, 198 people total): South West Africa People's Organisation (SWAPO) (94, 81.7*

*Netherlands (2 elections, 340 people total): People's Party for Freedom and Democracy (39, 19.7*

*New Zealand (11 elections, 1280 people total): National Party (539, 42.1*

*Nicaragua (2 elections, 251 people total): Sandinista National Liberation Front (75, 54.0*

*Niger (2 elections, 377 people total): Nigerien Party for Democracy and Socialism (PNDS Tarayya) (74, 40.4*

*Nigeria (2 elections, 597 people total): People's Democratic Party (132, 83.5*

*Norway (2 elections, 383 people total): Norwegian Labour Party (102, 38.8*

*Pakistan (2 elections, 701 people total): Pakistan Muslim League (Nawaz faction) (222, 32.7*

*Palestinian Territory, Occupied (2 elections, 264 people total): Fatah (115, 44.1*

*Panama (3 elections, 214 people total): Democratic Change (82, 38.5*

*Paraguay (2 elections, 286 people total): National Republican Association (Colorado Party) (137, 48.9*

*Peru (2 elections, 320 people total): Peruvian Aprista Party (85, 26.9*

*Philippines (2 elections, 685 people total): Liberal Party (160, 23.6*

*Poland (2 elections, 986 people total): Law and Justice (424, 43.5*

*Portugal (3 elections, 822 people total): Socialist Party (379, 46.4*

*Puerto Rico (2 elections, 161 people total): New Progressive Party (PNP) (112, 70.4*

*Qatar (2 elections, 114 people total): Independent (3, 37.5*

*Romania (2 elections, 703 people total): Social Democratic Party (215, 31.1*

*Russian Federation (8 elections, 3591 people total): United Russia (1441, 40.1*

*Rwanda (2 elections, 289 people total): Rwandan Patriotic Front (RPF-Inkotanyi) (214, 74.0*

*Saudi Arabia (2 elections, 396 people total): N/A (12, 60.0*

*Senegal (2 elections, 198 people total): Senegalese Democratic Party (162, 81.8*

*Serbia (2 elections, 607 people total): Serbian Progressive Party (SNS) (124, 20.7*

*Sierra Leone (2 elections, 438 people total): All People's Congress (250, 57.6*

*Singapore (2 elections, 231 people total): People's Action Party (211, 93.8*

*Slovakia (2 elections, 349 people total): SMER - Social Democracy (119, 34.1*

*Slovenia (2 elections, 235 people total): Slovenian Democratic Party (62, 26.6*

*Solomon Islands (2 elections, 125 people total): Association of Independent Members (AIM) Party (37, 30.1*

*Somalia (2 elections, 288 people total): Independent (8, 29.6*

*Somaliland (1 elections, 113 people total): Peace, Unity and Development Party ( Kulmiye) (57, 50.4*

*South Africa (2 elections, 847 people total): African National Congress (ANC) (542, 64.8*

*South Korea (2 elections, 627 people total): Democratic Party (256, 42.2*

*Spain (2 elections, 831 people total): Popular Party (321, 38.9*

*Sri Lanka (2 elections, 456 people total): Sri Lanka Podujana Peramuna (SLPP) (45, 14.8*

*Suriname (1 elections, 71 people total): NDP (13, 26.5*

*Sweden (2 elections, 765 people total): Social Democrats (240, 31.8*

*Switzerland (2 elections, 523 people total): Swiss People's Party SVP (141, 27.2*

*Syria (Syrian Arab Republic) (2 elections, 405 people total): Arab Socialist Baath Party (365,*

91.5

*Taiwan (2 elections, 225 people total): DPP (131, 58.7*

*Tajikistan (2 elections, 184 people total): People's Democratic Party of Tajikistan (159, 88.3*

*Thailand (4 elections, 2000 people total): Thai Rak Thai (632, 31.6*

*Timor-Leste (East Timor) (2 elections, 233 people total): National Council of Timorese Reconstruction (CNRT) (96, 41.7*

*Trinidad and Tobago (2 elections, 108 people total): United National Congress (62, 57.4*

*Tunisia (2 elections, 486 people total): Democratic Forum for Labour and Liberties (200, 41.3*

*Turkey (2 elections, 1131 people total): Justice and Development Party (643, 57.4*

*Turkmenistan (2 elections, 392 people total): Democratic Party of Turkmenistan (391, 100.0*

*Ukraine (3 elections, 1349 people total): Servant of the People (254, 19.2*

*United Arab Emirates (2 elections, 77 people total): None (77, 100.0*

*United Kingdom (Great Britain) (2 elections, 1283 people total): The Conservative Party (615, 48.2*

*United States of America (2 elections, 981 people total): Republican Party (530, 54.2*

*Uruguay (2 elections, 274 people total): Broad Front (143, 53.0*

*Uzbekistan (2 elections, 352 people total): The Liberal Democratic Party of Uzbekistan (UzLiDeP) (208, 59.9*

*Venezuela, Bolivarian Republic of (2 elections, 394 people total): United Socialist Party of Venezuela (180, 46.0*

*Viet Nam (2 elections, 971 people total): Communist Party of Vietnam (962, 99.9*

*Yemen (2 elections, 688 people total): General Peoples Congress (412, 60.0*

*Zambia (2 elections, 372 people total): Patriotic Front (181, 49.2*

*Zimbabwe (2 elections, 477 people total): Zimbabwe African National Union-Patriotic Front (263, 56.2*

## 9 Appendix D

### 9.1 AI Agent Details

We deploy a retrieval-augmented AI search agent (ASA) to infer missing attributes, here, party affiliation for a political leader in a particular year, from publicly available digital sources. The agent operates under a country-specific, closed codebook provided by our experts and may either assign a single label from that set or abstain from doing so.

As visualized in Figure 13, for each MP the agent queries parliamentary portals, official biographies, reputable news, and encyclopedic entries; extracts corroborating statements; and returns a structured prediction (label, one-sentence rationale, and source citations). The agent is instructed to prioritize precision over recall and to abstain when evidence is weak or conflicting sources are present. All queries, snippets, and links are archived to ensure auditability and replication. This provenance-preserving design expands coverage while limiting measurement error and maintaining cross-national comparability.

**Design goals.** The verifiable AI search agent (ASA) is engineered to expand coverage of missing labels while preserving cross-national comparability and auditability. Three principles guide the design: (i) *precision over recall* via conservative prediction and a confidence gate that withholds uncertain cases; (ii) *closed-world classification* against country-specific codebooks supplied by experts; and (iii) *provenance preservation*, archiving all queries, snippets, and Internet links for replication.

**System architecture.** The ASA is a retrieval-augmented ReAct agent implemented with `langgraph`. A small instruction-tuned LLM (GPT5-nano, `temperature` = 1) orchestrates tool use and produces a one-shot JSON output containing a label, a one-sentence rationale with citations, and a confidence flag (High/Medium/Low).

The agent has access to two read-only connectors: a Wikipedia retriever and a general web search tool (DuckDuckGo). All decisions are constrained to a *closed codebook* (per country) delivered in-prompt; the model is explicitly instructed to abstain from guessing and to prioritize verifiable sources over known or implicit heuristics.

**Data flow and reproducibility.** Input candidate records (GLP leaders) are materialized in `SQLite`. For each row, the agent executes a short search session and returns a structured prediction. Results—including raw tool traces—are appended to a `SQLite` store. Post-hoc analysis, scoring, and figure generation are performed in `R`.

**Country-closed matching and normalization.** To guard comparability across countries and reduce typographic drift, the pipeline applies a two-stage, codebook-guided normalization to the model’s raw label string:

1. **Strict closed-set match (country scope).** If the predicted string exactly matches a member of the country’s codebook, it is accepted.
2. **Conservative fuzzy match.** Otherwise, a relaxed mapper computes a similarity score  $s$  combining (a) Jaro–Winkler similarity on a normalized label, (b) token overlap coverage, and (c) acronym equality. Let  $m$  denote the runner-up score. Accept as a match if if

$$s \geq 0.92 \quad \text{or} \quad (s \geq 0.85 \ \& \ s - m \geq 0.08).$$

This mapping process produces a normalized label used for evaluation and aggregation while preserving the original string for audit. It corrects innocuous variants (pluralization, punctuation, acronyms) without introducing new classes. The mapper's conservatism is intentional: tightening thresholds reduces false positives at the cost of additional abstentions.

**Confidence gating and abstention policy.** The agent emits a categorical confidence estimate; records flagged Low or Medium are withheld from downstream analyses. This implements an *abstention* layer that trades coverage for precision and is especially important where web evidence is sparse, parties are newly formed, or transliterations vary. As documented in the main text, low-confidence rates vary by country and cohort and are higher for small/minority parties, earlier periods, and cases lacking standardized biographies.

**Task prompt.** For the party prediction task, see below for the prompt sent to the AI search agent:

```

TASK OVERVIEW:
You are an advanced search-enabled Large Language Model (LLM) specializing in party affiliation inference.
Your role is to determine the most likely political party of well-known political leaders or other notable individuals,
strictly from a provided list of possible parties. Your answers are based on:

1. Publicly available background data and official information (obtained via search tools),
2. Contextual evidence such as party membership records, voting history, public statements, or credible news sources.

ACCESS TO SEARCH:
You are REQUIRED to first use query search tools to research the name in question.
Search for any authoritative or credible sources referencing the individual's
official party membership or widely recognized affiliation.
If you find clear, credible information on the person's party, rely on it.
If information is conflicting or indeterminate, then fall back on contextual inference
grounded in legislative records, news coverage, or the individual's own statements.

TARGET INDIVIDUAL:
* Name: <PERSON-NAME>
* Country: <COUNTRY>
* Approximate year: <YEAR>
* Potential Parties in this Country (PARTIES\_OF\_COUNTRY): {<PARTIES-OF-COUNTRY>}

CONSTRAINTS FOR "pol\_party":
1. You MUST choose exactly ONE party from the above list for "pol\_party".
2. You must NOT introduce any party that is not in the list.
3. You must preserve EXACT spelling, capitalization, and punctuation
   for the chosen party as it appears in the list (INCLUDING ABBREVIATIONS).
4. All explanations should be written in English.

RESPONSE FORMAT:
Your output must follow this precise JSON structure (and nothing else):
{
  "justification": "A concise one-sentence justification citing either the external source findings or,
   if no consensus, contextual inference from legislative or news records.",
  "pol\_party": "A party from PARTIES\_OF\_COUNTRY with exact same spelling, capitalization/punctuation/abbreviation style.",
  "pol\_party\_relaxed": "One party, best prediction, no constraints.",
  "confidence": "Confidence in your answer (High, Medium or Low)."
}

IMPORTANT REQUIREMENTS:
* Do NOT include additional text or commentary beyond the JSON object.
* If you find verifiable sources confirming the individual's party, reference them EXPLICITLY
   within your single-sentence justification.
* If no definitive sources exist, clearly state that your choice is based on contextual inference.
* If there are remaining ambiguities, select the MOST likely
   based on searched content and public records.

FINAL TASK STEPS:
1. Use search tools to verify the individual's publicly acknowledged party membership.
2. If confirmed, select that party from the PARTIES\_OF\_COUNTRY list.
3. If conflicting or no direct sources, apply best-effort sources-first, context-second inference.
4. Output the strict JSON block:
{
  "justification": "...",
  "pol\_party": "...",
  "pol\_party\_relaxed": "...",
  "confidence": "..."
}

WARNINGS:
* Under NO circumstances produce any output outside the JSON format.
* Any deviation from this exact JSON structure risks rejection.

```

\* Justification must be ONE sentence only.  
\* Party must match EXACTLY the spelling in the list.

**Hyperparameters and their consequences.** Key choices and expected effects are as follows:

1. **LLM decoding.** A near-deterministic setting reduces variance in structured outputs and improves JSON validity, at the cost of slightly lower recall on ambiguous cases.
2. **Fuzzy-match thresholds.** Country thresholds (0.92; , 0.85+0.08) and stricter global thresholds (0.95; , 0.90+0.12) suppress false alignments across near-synonyms, benefiting cross-national comparability while increasing abstentions where parties have overlapping or evolving brands.
3. **Confidence gate.** Withholding Low and Medium confidence predictions curbs label noise—especially for minority and emergent parties—but mechanically reduces coverage; this is the intended precision-first behavior for a reference-quality dataset.

**Languages and programs.** The agentic layer is implemented in **Python** with `langgraph` (React) and OpenAI-compatible chat bindings; retrieval uses Wikipedia and general web search connectors. Orchestration is done in R. Caching and data interchange use **SQLite** with `DBI/RSQLite`.

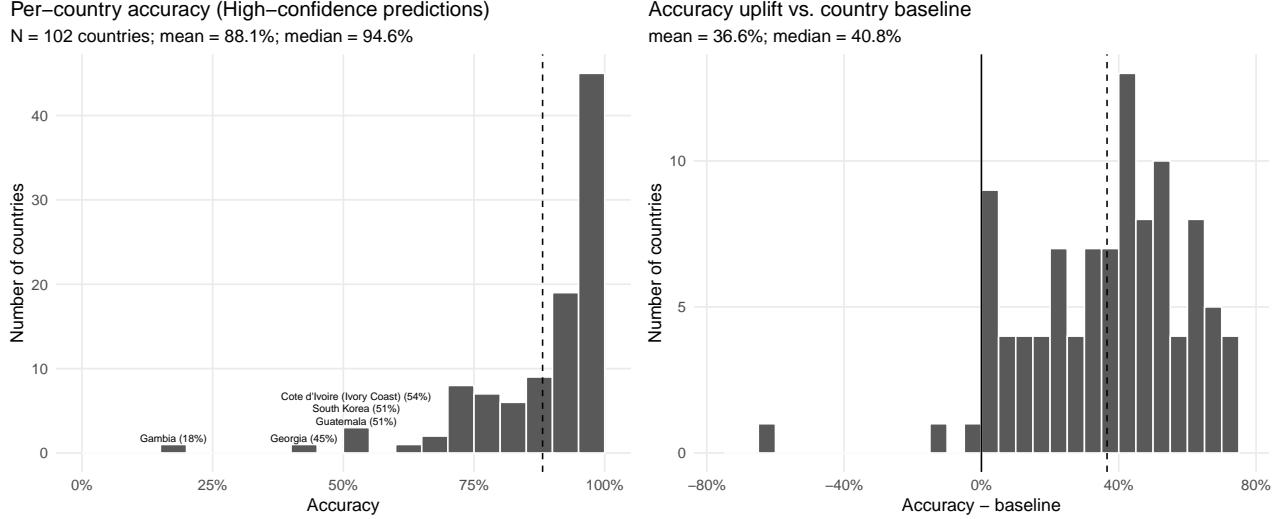
**Applicability and limitations.** The design is most reliable for *verifiable attributes* (e.g., party membership), where authoritative sources exist and closed codebooks are stable. Performance is predictably uneven for (i) very small or new parties, (ii) early cohorts with sparse web footprints, and (iii) languages/alphabets with heterogeneous transliteration conventions. The relaxed mapper mitigates harmless surface variation without altering class boundaries, and the abstention gate limits error propagation. While the same architecture extends to other attributes (e.g., ethnicity, religion), attributes lacking public self-identification should be treated with additional caution and stricter confidence gating.

**Agent Performance.** To gauge the performance of this protocol we focus on missing party affiliations rather than ethnicity because it is objectively verifiable through public records and exhibits substantially higher rates of missing data, providing a more robust test.

Using the subset of records for which expert codes are available, this verifiable search agent attains high agreement on party labels. Across  $N = 114$  countries and 34,618 matched leader-records covering 1,209 distinct party labels, the overall high-confidence accuracy is 0.860, with a mean per-country high-confidence accuracy of 0.879 (Figure 11). Relative to a simple country-majority baseline of 0.536, this yields an average uplift of 0.343 points on a 0–1 scale, ranging from -0.602 in Gambia to 0.761 in Finland. Point estimates improve in later election cohorts, consistent with expanding web coverage and corpus quality (Figure 12). If we ignore the confidence gate and score all predictions, overall accuracy is 0.751.

Limitations, however, remain. On average across countries, a share of 0.250 of records is flagged “Low” and withheld from downstream analysis, yielding a conservative working sample. Performance is predictably uneven across label frequency: mean high-confidence accuracy for small/minority party labels is 0.692, compared to 0.820 for plurality parties. In practice, disagreements and abstentions concentrate in (i) very small or newly formed parties, (ii) earlier election years, and (iii) cases with sparse or non-standardized online biographies.

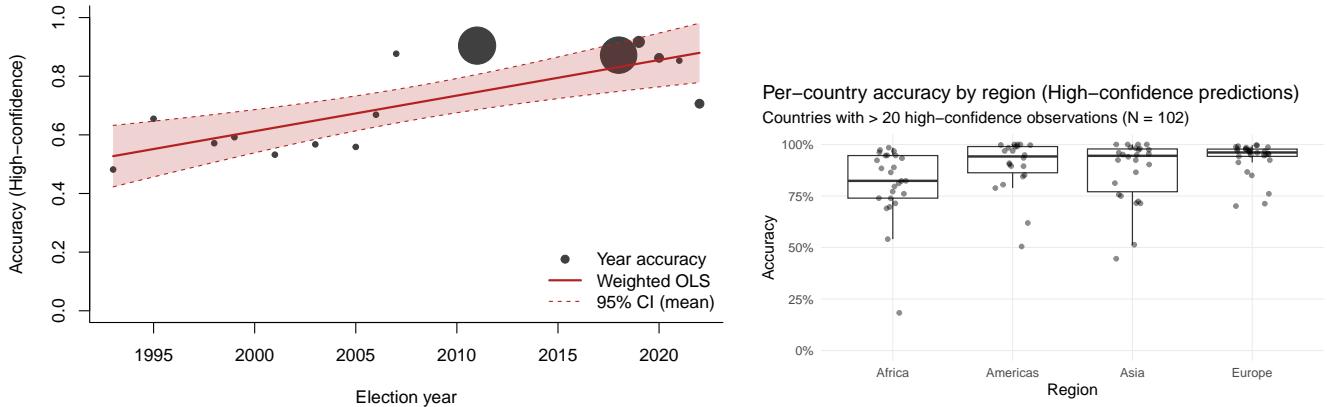
Overall, this provenance-preserving search agent substantially expands usable coverage by 20.8% (12,898 observations with previously missing party labels) at low marginal cost while maintaining high out-of-sample agreement with expert codings (Figure 11). The approach is transparent in that



**Figure 11:** Agent performance in predicting party, across countries having at least 20 agent codes. Median country-level performance is above 90%. The baseline refers to the accuracy of predicting that every politician in a given country takes on the dominant characteristic being considered.

every assigned label is backed by stored queries and citations—portable across countries because predictions are constrained by closed codebooks. Our downstream estimates will employ the high-confidence entries from the agent. For results using only expert coding, see Appendix A; for additional details about the agent architecture, see Appendix B. See Table 10 for a sample of the sources obtained by the agent.

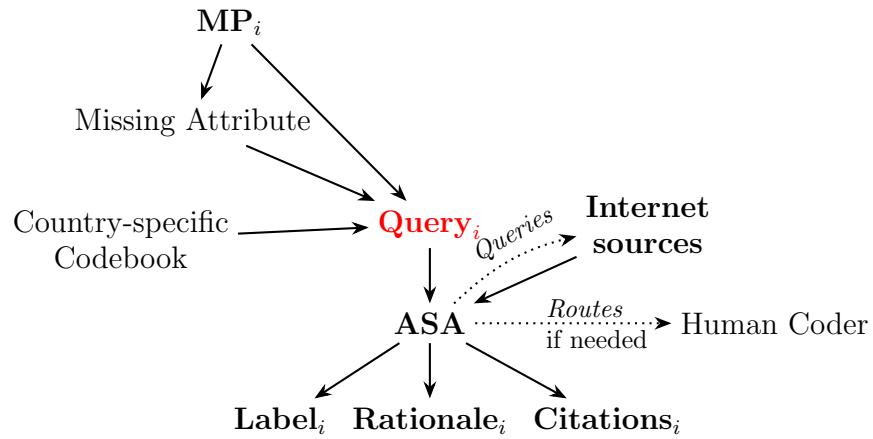
**Summary.** In sum, the ASA combines a *langgraph* ReAct controller, source-first prompts, conservative similarity thresholds, and a confidence gate to deliver auditable, high-precision labels at scale. The resulting pipeline is portable across countries (closed codebooks), reproducible (stored traces and caches), and well-suited to augment expert codings with verifiable, citation-backed inferences.



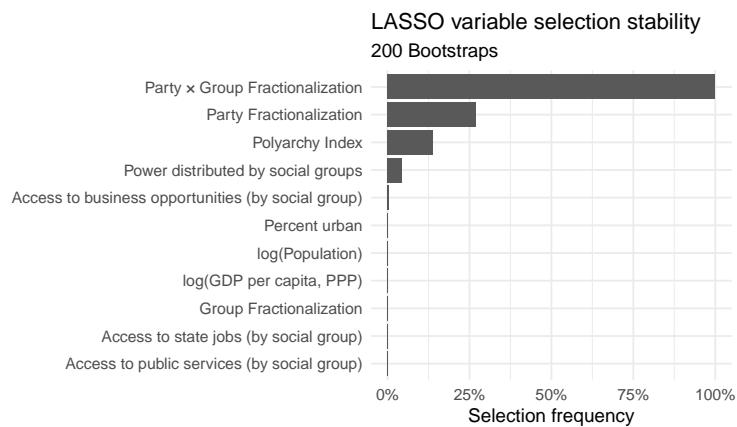
**Figure 12:** Agent performance over time; point size corresponds to the number of observations in each (approximate) year (LEFT) and space (RIGHT).

**Table 10:** Sample agent traces. Text content truncated for readability (and may contain typographical errors as present in native source). Links are clickable. Full traces contain many more sources.

Field	Content
<b>Entry 1: Syleiman Abusaidovich Kerimov — Russian Federation (1999)</b>	
Country	Russian Federation
Year	1999
Person	Syleiman Abusaidovich Kerimov
Wikipedia	Page: Ashot Egiazaryan Summary: Ashot Gevorkovich Egiazaryan (Russian: ; Armenian: ; born...)
Search 1	In the spring of 1998, Yeltsin dismissed Chernomyrdin as head of government and in 1999 Yeltsin's administration backed a newly formed party, Un...
URL 1	<a href="https://en.wikipedia.org/">https://en.wikipedia.org/...</a>
Search 2	OURHOMEISRUSSIAPARTYOurHomeIsRussia(Nash Dom—Rossiya, or NDR) was a sociopolitical movement and a ruling party from 1996 to 1998. Source for i...
URL 2	<a href="https://www.encyclopedia.com/">https://www.encyclopedia.com/...</a>
<b>Entry 2: Jasminka Stanojevic — Serbia (2018)</b>	
Country	Serbia
Year	2018
Person	Jasminka Stanojevic
Wikipedia	Page: Supreme Court (Serbia) Summary: The Supreme Court (Serbian: , romanized: Vrhovni sud) is the court of last resort in Serbia...
Search 1	This article lists political parties in Serbia, including parties that existed in the Kingdom of Serbia between the early 1860s and 1918. A kol...
URL 1	<a href="https://en.wikipedia.org/">https://en.wikipedia.org/...</a>
Search 2	Imali su dve i četiri godine kad smo izbegli iz Knina. Kad bi neko pokucao na vrata, vikali bi: „Tata, tata“. Tri godine nakon progona sazna...
URL 2	<a href="https://www.kurir.rs/">https://www.kurir.rs/...</a>
<b>Entry 3: Mihai STROE — Romania (2011)</b>	
Country	Romania
Year	2011
Person	Mihai STROE
Wikipedia	Page: Adrian Stroe Summary: Adrian Stroe (born 24 October 1959), known as The Taxi Driver of Death, is a Romanian serial killer responsible ...
Search 1	Născut în Bucureşti și cu origini în comuna argeșeană Morăreşti, fost medaliat cu aur la olimpiada internațională de informatică, Mihai Stroe(...)
URL 1	<a href="https://adevarul.ro/">https://adevarul.ro/...</a>
Search 2	Mihai STROE Parliamentary activity in legislature 2008–2012 DEPUTY Constituency no.38 TULCEA, uninominal college no.2 Member al PDL, deputatul...
URL 2	<a href="https://www.cdep.ro/">https://www.cdep.ro/...</a>
<b>Entry 4: Matsie Angelina Motshekga — South Africa (2018)</b>	
Country	South Africa
Year	2018
Person	Matsie Angelina Motshekga
Wikipedia	Page: Angie Motshekga Summary: Matsie Angelina "Angie" Motshekga (born 19 June 1955) is a South African politician and educator who is currently...
Search 1	Matsie Angelina "Angie" Motshekga (born 19 June 1955) is a South African politician and educator who is currently serving as the Minister of Defense...
URL 1	<a href="https://en.wikipedia.org/wiki/Angie_Motshekga">https://en.wikipedia.org/wiki/Angie_Motshekga</a>
Search 2	Motshekga was elected the national president of the African National Congress Women's League (ANCWL) in 2008, defeating the League's secretary-general...
URL 2	<a href="https://www.sahistory.org.za/people/matsie-angelina-motshekga-angie-motshekga">https://www.sahistory.org.za/people/matsie-angelina-motshekga-angie-motshekga</a>



**Figure 13:** AI Search Agent (ASA) visualization.



**Figure 14:** LASSO selection.